Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science & Statistics
ADAPT Research Centre
Knowledge and Data Engineering Group (KDEG)

Enhancing the Transparency of Personal Data Access through Semantic Web Technologies

Roghaiyeh(Ramisa) Gachpaz Hamed

Supervisors:
Prof. Owen Conlan
Prof. Declan O'Sullivan

A thesis submitted to the
University of Dublin, Trinity College
in fulfilment of the requirements of the degree of
Doctor of Philosophy

2021

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed:

Roghaiyeh(Ramisa) Gachpaz Hamed

Permission to lend or copy

I, the undersigned, agree that the Trinity College Library may lend or copy this thesis upon request.

Signed:

Roghaiyeh(Ramisa) Gachpaz Hamed

Date:

25.02.2021

Acknowledgements

First and foremost, I would like to express my sincere appreciation to my supervisors, Prof. Owen Conlan and Prof. Declan O'Sullivan, whose their invaluable advice, continuous support, patience and encouragement have made this work possible.

As importantly, I am extremely grateful to my family- Ali who has always been there for me taking away all the worries and giving me unconditional support, endless happiness and love; my parents whom I would never become who I am today if it was not for their support and inspiration, especially my mom, the first and most influential role model in my life, who has made me believe that I can achieve anything I want; and my brother, who was always there to give me an ear whenever I need someone to talk to.

I would like to thank the Science Foundation of Ireland for funding this PhD work and the ADAPT Centre for providing excellent academic infrastructure for conducting this research.

Also, I would like to express my gratitude to my friends and colleagues (past and present) in the School of Computer Science and Statistics (SCSS) at Trinity College, Dublin.

My gratitude extends to my friends in the ADAPT Centre, especially Harshvardhan Pandit, Annalina Caputo, Brendan Spillane, Gary Munnelly, Emma Clarke, Kieran Fraser, Kaniz Fatema, Retno Vinarti, Kris McGlinn, Jamie McGann, Christophe Debruyne, Sinead Impey, whom have helped me, one way or another, in the completion of this thesis.

Moreover, I wish to show my gratitude to Prof. Séamus Lawless for his inspiration, and Prof. Dave Lewis and Prof. Gaye Stephens for their generous advice.

Also, I wish to especially thank my friends; Nazi, Manizheh, Sara, Fatemeh and Nima, for their empathy and sympathy to cheer me up during the challenging times of this journey.

Finally, I wish to express my deepest gratitude to The God for making all things possible and helping me to carry out my PhD research.

Thank you all.

Abstract

Nowadays, as the application of data-driven technologies and their influence in our lives grow exponentially, the amount of users' information collected, stored, and exchanged increases accordingly. Therefore, it is practically impossible for individuals to keep track of all the traces of their information. Consequently, users have a concern about the protection of their personal data. On the other hand, if people make their data strictly private, it could be depriving them of all of the advantages and benefits of these online services and facilities. There is a tremendous advantage to users in sharing the right information with the right people in the right ways; scientists can use data unexpectedly and discover ground-breaking results that can cure diseases, predict disasters, improve human behaviour and facilitate their lives.

Access control mechanisms alone have been proven ineffective at addressing modern privacy problems, and transparency plays a crucial role in enriching individuals with control over their data by providing them with sufficient knowledge regarding their personal data processing and helping them make well-informed decisions at the moment of data disclosure.

Accordingly, worldwide data protection laws and regulations, such as the European General Data Protection Regulation (GDPR), explicitly include transparency rules to oblige data processing parties to reveal respective information to the data subjects. These obligations are typically fulfilled through respective transparency parts of written privacy policies. However, such privacy policies exhibit several shortcomings that severely limit their actual reception and comprehension on the side of data subjects: First of all, privacy policies are often long, complex, and written in legalese language, making it hard for data subjects to locate transparency-related information and understand them correctly. Second, different privacy policies employ different logical structures and vocabularies for factually similar statements, causing significant reading efforts for every new policy to be understood. These drawbacks lead to a state where privacy policies are not read anymore before using a particular service and consenting to a specific collection and use of personal data. Under such conditions, transparency statements increasingly degenerate into rather self-serving formal compliance exercises instead of supporting data subjects' informed decisions and privacy-preserving conduct.

This thesis introduces the conceptual design of a novel service, named "eXplainable Personal Data Access" (XPDA), to enhance the control of individuals over their personal data access by leveraging Semantic Web technologies. The service has adopted the best practice of the existing access control model to exploit context-awareness and policy specification. Meanwhile, the service enhances the transparency of the privacy rules implications on access decisions by revealing the data access, explaining its reason and representing all of this information in a way that individuals could understand. Finally, a prototypical implementation of this service on a motivating scenario in the health domain demonstrates its adequacy to fulfil all the above-mentioned design goals.

In this research, a comprehensive user study is designed to evaluate the extent to which non-expert people can perceive the practical advantage of an explanation generated through the XPDA. The user study experiment deploys a quantitative approach to assess three well-agreed concepts of measurement for evaluating the interpretability of generated explanations. Experimental design for evaluating the usability of explanations and satisfaction of users adopts standard questionnaires and approaches. Moreover, a novel method is proposed to design the experiment to assess the understandability of the explanations considering different aspects of understanding. Finally, the impact of different evaluation factors is investigated through the statistical analysis of the results. The user study results show that the XPDA service can generate sufficiently usable explanations perceived with a high level of understanding and satisfaction for most participants.

Therefore, the service proposed in this thesis can benefit data subjects to obtain their right to the protection of their personal data and allow them to avail their right to be informed about the collection and use of their personal data. Meanwhile, the research community can deploy and advance it in other domains, and data controllers and service providers could advance it for auditing and assessing personal data access.

Contents

List of Figures

List of Tables

List of Abbreviations & Acronyms

ABAC	Attribute-Based Access Control
ACU	Access Control Unit
ABox	Assertional Box
AE	Access Exposer
AI	Artificial intelligence
ANOVA	Analysis of Variances
ASQ	After-Scenario Questionnaire
ATU	Access Transparency Unit
AX	Access eXplainer
CBAC	Context-Based Access Control
CM	Context Manager
DAC	Discretionary Access Control
DAML	DARPA Agent Markup Language
DARPA	Defense Advanced Research Projects Agency
DPV	Data Privacy Vocabulary
GDPR	General Data Protection Regulation
HCI	Human - Computer Interaction
IML	Interpretable Machine Learning
IoT	Internet of Things
KAoS	Knowledgeable Agent-oriented System
KM	Knowledge Manager
KMU	Knowledge Modelling Unit
KR	ontological Knowledge Representer
MAC	Mandatory Access Control
MCQ	Multiple-Choice Question
OWL	Web Ontology Language
OWL-DL	OWL Description Logic
PEM	Policy Enforcement Manager
PET	Privacy Enhancing Technologies
RBAC	Role-Based Access Control

RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RE	Reasoning Engine
SABAC	Semantic-aware Attribute-Based Access Control model
SotA	State of the Art
SPARQL	Simple Protocol and RDF Query Language
SUS	System Usability Scale
SWRL	Semantic Web Rule Language
TBox	Terminological Box
TET	Transparency Enhancing Technologies
UML	Unified Modelling Language
XACML	eXtensible Access Control Markup Language
XAI	eXplainable Artificial Intelligence
XG	eXplanation Generator
XI	eXplanation Interpreter
XML	Extensible Markup Language
XPDA	eXplainable Personal Data Access

1 Introduction

1.1 Motivation

On the World Wide Web's 28th birthday in March 2017, Sir Tim Berners-Lee claimed that *losing control over our personal data* is one of the three biggest challenges facing the web today [1].

Daily growth in the use of new digital devices, technologies and services has given rise to various ways to collect, access, analyse and use personal data. While online service users often share their data consciously and voluntarily, in many other instances, data are collected without their knowledge. This circumstance is defined as "information asymmetry", where the "data about us are collected in the circumstances we may not understand, for purposes we may not understand and are used in ways we may not understand" [2]. Consequently, this lack of awareness endangers users' privacy [3] [4] and raises severe privacy concerns. [5] [6] [7] [8] [9] [10].

To address privacy concerns in the new digital age, the traditional privacy paradigm of concealment (i.e. controlling the access to and distribution of personal data) does no longer hold or is impossible to maintain [11] [12]. Nowadays, few online services do not collect, access, or use the online data of users to provide more beneficial services to them. Therefore, new privacy-preserving approaches should afford users better control over their data usage [13]. This control entails providing people with knowledge of:

- what information is collected about them,
- what/who are these third parties which get access to this information and
- why this information is transferred onward or made available to third parties.

Current studies emphasise that people are willing to disclose their private information (even information that allows them to be personally identified) when perceiving themselves to be in control over the release and access of these data [14] [15] [16] [17] [18]. Meanwhile, the lack of control over their data may be preventing people from getting to grips with the internet and could perpetuate the "digital divide", with many people missing out on online engagement benefits [19].

A prerequisite for a high level of people's control over their personal data is the transparency on personal data processing [17] [20]. Transparency can be considered people's ability to obtain "an adequate level of clarity of the processes in privacy-relevant data processing" [21]. Therefore, it would allow people to have better control over their data if procedures of the collection, storage, usage, and removal of their personal data are revealed in a clear, transparent and reversible manner [22] [23] [24].

The need for transparency of data gathering and usage is also emphasised in the General Data Protection Regulation (GDPR) [25], to allow people to exercise their right to protect their personal data. Articles 13 and 14 give EU citizens a right to be informed about their personal data collection and use. Article 15 provides the data subject with rights to attain "meaningful information about the logic involved" in automated decisions. GDPR defines 'data subject' as "identified or identifiable natural person[s]" [25]. In other words, the data subject is referred to an individual whose data is collected/processed. Recital 71, which supplements Article 22 as a whole, additionally states more safeguards including specific information to the data subject, and a right to "obtain an explanation of the decision reached after such assessment". At the same time, companies are also required to comply with several fundamental principles set out in Article 5 in order to be able to process personal data. These obligations include but are not limited to, processing personal data in a lawful, fair, and transparent manner to the data subject.

While transparency is necessary to provide information available and accessible to people, it is not sufficient on its own. The information needs to be understood and comprehended by their receivers to achieve useful transparency [26]. It is pointed out that "transparency can only be useful when it enhances the understanding, not just increasing the flow of information" [27]. Regulations also enforce to make transparency more understandable to the public. Article 12 of the GDPR and accordingly Recitals 39 and 59, have declared that any transparency information related to data processing should be provided to the data subject in a "concise, easily accessible form". It should be "intelligible" and "easy to understand" and should be provided "using clear and plain language". The meanings and definitions of these terms are open to interpretation and differ across different studies of various scientific communities. Therefore, it is essential to know these interpretations, understand their differences and identify a set of features that can be considered as a concept of measurement for transparency.

Another consideration which can also affect achieving useful transparency is the perception of people and their understanding. Transparency may not be accomplished if provided information does not match with the comprehension level of people. Since it is usually difficult to match people's perception, it has been recommended to find some information interpretation methods and representation, each of which suits different group of people [28]. Furthermore, since perception is subjective, various individuals perceive the same information differently and respond to the information diversely according to their perception [29]. Therefore, in order to enhance transparency, the most straightforward representation of information should be chosen to match a broad range of people's perception. Alternatively, people should be able to choose a representation of information which maximises their understanding.

All of these studies indicate the need for further investigation on identifying the new approaches to give people more control over their personal data. These approaches need to control access and usage of data by providing and enhancing the transparency in a comprehensible way.

1.2 Research Question

The research question investigated in this thesis is:

Research Question

To what extent can a Semantic Web-based service enhance transparency to a human on her/his personal data access in an interpretable manner?

In this study, *Semantic Web-based service* is referred to a service which can deploy the collection of Semantic Web technologies to:

- Describe, represent and organise the knowledge, by defining main concepts and their relationships (defining vocabularies and ontologies)
- Retrieve, extract and expose these relations' information by providing one or more patterns against such relations (querying)
- Analyse the content of the data to discover new relationships and possible inconsistencies (inferring)

Although it will be discussed in further detail in Section 2.4, throughout this thesis, *transparency* would be conceptualised in the sense of clarity of information as described in [30] and would be referred as proven knowledge about data that have been disclosed previously. Therefore, this research focuses on the aspect of *transparency* as the ability of a data subject to obtain "an adequate level of clarity of the processes in privacy-relevant data processing" [21].

Also, derived from [25], *personal data* is defined as "any information relating to an identified or identifiable natural person (data subject); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”.

Moreover, *interpretability* is defined as the ability “to explain or tell the meaning in understandable terms” in [31]. This definition is considered as a system-centric definition of *interpretability* in [32]. The same study defined interpretability as “the degree to which an observer can understand the cause of a decision” according to its human-centred perspective [32]. Both system-centric and human-centric definitions of interpretability would be considered in this thesis.

1.2.1 Research Objectives

This section describes the main objectives of this research in order to address the research question outlined above.

The first research objective focuses on the literature review to explore the related works. The second research objective is structured on exploiting Semantic Web technologies to provide data subjects with enhancing transparency on their personal data access in an interpretable manner. The third objective concentrates on implementing this service in a motivating scenario to provide more control over personal data access by revealing the access and clarifying its reason to be understood by an individual. Finally, the last objective is associated with assessing the represented explanation.

Firstly, literature review of state of the art needs to be conducted to synthesis previous research, identify gaps of the current state of knowledge and justify the research question, which provides the first objective as:

***RO*1:** To review the literature on:

- the use of Semantic Web technologies to advance characteristics of Access Control, and
- the specifications of Transparency Enhancement Technologies.

To propose a service for extending control over data access through enhancing transparency, the detail of its architecture needs to be identified and designed according to the functionality of service components. These steps can be stated as the second objective:

***RO*2:** To design a service to exploit Semantic Web technologies to provide transparency on personal data access in an interpretable manner.

The designed service needs to be implemented as a prototype on a simple arbitrary scenario of personal data access to prove that all different components of its architecture work appropriately as an end-to-end service and all expected functionality of the service can be achieved. This requirement highlights the third objective as:

***RO*3:** To implement an end-to-end prototype of service in a sample scenario to provide more control over personal data access by revealing the access and clarifying its reason in a way that can be understood by an individual.

It needs to be assessed the extent to which the provided service can fulfil the expected functionalities. Designing an evaluation method based on standard methodologies and defining metrics and measuring methods to assess the system outputs triggers the fourth objective as:

***RO4*:** To evaluate the interpretability of outputs provided by the service.

1.3 Research Methodology

1.3.1 Literature review

As an initial step, related literature were reviewed to achieve the appropriate knowledge about different features of access control approaches. The inclusion of these approaches mainly focused on the deployment of Semantic Web technologies and their capabilities towards addressing these features.

Besides, we explored different research conducted to provide human-centred and user-friendly transparency of personal data access. This investigation helped us to find out different criteria for transparency in personal data access. The result of this exploration is represented in Chapter 2 and fulfilled RO1.

1.3.2 Architecture for the proposed service

To fulfil RO2, and according to outcomes of the SotA review, a need for developing a service to facilitate understating of data subjects on their personal data access was recognised. Therefore, a novel service exploiting the Semantic Web technologies, called "eXplainable Personal Data Access" (XPDA), was designed. The architecture was modelled and documented using the graphical notation of C4 model[1], which considers the static structures of a software system through visualising a hierarchy of abstractions.

The architecture adopts context-based access control model for policy adaptation. Another fundamental property of access control model in XPDA is deploying the deductive capabilities of an ontological approach along with run-time inference capabilities of a rule-based approach. Combination of above approaches enables the efficient enforcement of policies defined over dynamically determined context values.

Moreover, the architecture of XPDA aims to provide data subjects with ex-post transparency through the understandable insights about their personal data access and conformance of the access with the policies. These insights are explained in a manner in which data subjects can understand the detail and cause of a data access decision.

1.3.3 Service development and implementation

The service was developed corresponding to the proposed architecture. Semantic Web technologies were deployed in order to implement the functionalities of each component of the architecture in a motivating scenario in the health domain. The detailed description of various steps required for implementing the service and its different outputs fulfilled RO3. Although details of the development steps are described in Chapter 4, a summary of each step can be found as follow:

- **Knowledge Modeling**

As the first step in knowledge modelling, an ontology was developed to represent the structure of data access by pointing out a set of related concepts and their properties and the relations between them in the motivating scenario. To develop this ontology, commonly adopted and recommended methodologies within the Semantic Web

[1] https://c4model.com/

technologies community were deployed, and the ontology was specified and represented by Web Ontology Language (OWL) [33].

Meanwhile, different circumstances of privacy rule specification and enforcement on various kind of represented knowledge of the motivating scenario were defined through three different use cases. The data access rules corresponding to each use case were defined using Semantic Web Rule Language (SWRL) [34]. Details of knowledge modelling are discussed in Section 4.2.

- **Access control and decision making**

Once the development and instantiation of the ontology are done, the model becomes a shareable explicit knowledge that could be considered as a repository. Before the repository could be accessed through different queries, additional implicit facts need to be inferred from explicit facts in the model and expand the knowledge. Inference and reasoning are mechanisms to discover additional information that is not explicitly stated in the initial data. Pellet reasoner [35] was used in the prototype to implement inference. The access decisions were then identified due to applying an appropriate query, written in Simple Protocol and RDF (Resource Description Framework) Query Language (SPARQL), to the latest inferred model. Details of reasoning and retrieving inferred information are discussed in Section 4.3.

- **Explaining the reason for the data access**

In order to explain the access decision, it requires to be figured out what has been stated in the ontology, which causes the decision. OWL Explanation API [36] was adopted to capture a minimal subset of the ontology sufficient for generating the inferred decision. It was also included corresponding access rule/s.

The explanation generated through OWL Explanation API needs to be converted into a format which can be perceived and comprehended by data subjects. Plain text as a most natural manner of communication and graph visualisation as a popular visual representation were used to express the explanation. Details of explanation and representation are discussed in Section 4.4.

1.3.4 Evaluation methodology

Evaluation of the proposed service was undertaken through a quantitative user study upon the implemented prototype. A group of participants with different knowledge backgrounds were recruited through Prolific [37] to assess the XPDA service outputs across three different use cases. A set of online questionnaires were designed and distributed using Qualtrics online survey platform [38].

An experiment was conducted to address RO4 to evaluate the extent to which data subjects can understand the explanation generated through the XPDA service about personal data access.

Three concepts of measurement, including usability, understandability and satisfaction, were measured across the experiment to assess the interpretability of the explanations. Chapter 5 describes the details of these evaluations with a brief review as follows:

- Usability of the XPDA service to expose and explain personal data access was tested using the System Usability Scale (SUS) [39] questionnaires. Results of the evaluation were analysed using different interpretations of SUS to ensure their validity.
- A set of questions was designed to evaluate the extent to which participants can understand the explanation generated through the XPDA service. Participants' responses for each question were scored using a conventional method. The results were analysed for within use cases and between use cases.
- Subsequently, participants' satisfaction on the explanation was also tested using After-Scenario Questionnaire (ASQ) [40]. After collecting all responses from participants, the ASQ scores were calculated and analysed.

1.4 Contributions

The major contribution of this thesis is the proposal of a novel service to provide data subjects with more transparency by generating an explanation on their data access and representing it in comprehensible formats. Another contribution of this study is to design a comprehensive user study to evaluate the interpretability of the generated explanation.

1.4.1 Design and development of the XPDA service

Designing a service and its prototypical implementation in order to explain personal data access to data subjects in an interpretable way is the major contribution of this thesis. The service leverages Semantic Web technologies to enhance the control of the data subject over their personal data access. Best practice of existing access control models has been adopted by involving careful consideration on exploiting context awareness and policy specification. Meanwhile, the service advances the state of the art by offering the approach to enhance the visibility on the implication of privacy rules on access decisions by providing detailed information about data access and explaining its justification. All of this information is represented in a way which could be perceived by non-expert users.

This service not only can benefit data subjects to obtain their right to the protection of their personal data but also allow them to avail their right to be informed about the collection and use of their personal data. Meanwhile, the research community can deploy and advance it in other domains, and data controllers and service providers could advance it for auditing and assessing the access over personal data.

1.4.2 Design a user study to evaluate the interpretability of the explanation

In this research, a comprehensive user study was designed to evaluate the extent to which participants can perceive the practical advantage of explanation generated through XPDA. The experiment of this user study deployed a quantitative approach to evaluate three well-agreed concepts of measurement for assessing the interpretability of generated explanation through XPDA. While standard questionnaires and approaches were adopted to evaluate two of them, i.e. usability and satisfaction, a very novel approach was used to design the experiment to evaluate the understandability of the explanation. This novel method was designed considering different aspects of understanding, including explicit, implicit and compositional cognitive chunks of the explanation [179]. The statistical analysis of the result also applied to more investigation on the impact of different factors of the evaluation.

1.4.3 Publications

The publications associated with the research in this thesis to date are:

- **"Semantic Reasoning for Privacy-Preserving Personalisation"** [41]
  R. G. Hamed, K. Fatema, O. Conlan, D. O'Sullivan
  11th International IFIP Summer School on Privacy and Identity Management, 2016.
  This paper, associated with RO2, presents the first iteration of the proposed service to focus on its high-level functionalities. Different required components are discussed in this paper by providing a primary use case.

- **"Explaining Disclosure Decisions Over Personal Data"** [42]
  R. G. Hamed, H. J. Pandit, D. O'Sullivan, O. Conlan
  18th International Semantic Web Conference (ISWC), 2019.
  This paper presents the earliest outcome of this research and demonstrates the implementation of our proposed service through a prototype on a sample use case. It consists of the earliest result of the user experiment as well. Therefore, its outcomes are relevant to RO2, RO3 and RO4.

Although the following publications were not directly related to the main focus of this thesis but were conducted by the author in the closely related research area including the application of Semantic Web technologies and the design of user-centric experiments.

- **"Creating a Vocabulary for Data Privacy"** [43]
  H. J. Pandit, A. Polleres, B. Bos, R. Brennan, B. Bruegger, F. J. Ekaputra, J. D. Fernández, R. G. Hamed, E. Kiesling, M. Lizar, E. Schlehahn, S. Steyskal, R. Wenning
  18th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE), 2019.
  Lack of agreed-upon vocabularies or taxonomies for describing personal data processing purposes and its categories is recognised across the XPDA service implementation and proposed to DPVCG[2]. This paper presents a collaborative

[2] https://www.w3.org/community/dpvcg/

work of the author of this thesis on the creation of Data Privacy Vocabulary (DPV) in order to address this gap by providing a comprehensive, standardised set of terms for annotating privacy policies, consent receipts, and in general records of personal data processing.

- **"The Use of Open Data to Improve the Repeatability of Adaptivity and Personalisation Experiment"** [44]
  H. J. Pandit, R. G. Hamed, S. Lawless, D. Lewis
  24th Conference on User Modeling, Adaptation and Personalization, 2016.
  This paper discusses how Semantic Web ontologies can be applied to the description and data of published adaptivity and personalisation experiments in a manner that can be linked from publications and easily located, accessed and reused to repeat an experiment.

- **"A Review of User-centred Information Retrieval Tasks"** [45]
  A. H.Vahid, R. G. Hamed, K. Koidl
  24th Conference on User Modeling, Adaptation and Personalization, 2016.
  This paper discussed the way of gathering users' profiles and objects of their interest in IR evaluation campaigns.

2 State of the Art

2.1 Introduction

This chapter describes the synthesis of previous research studies through a literature review to map and assess the research area, identify gaps to motivate this research, and justify the research question and hypotheses.

This literature review is conducted to identify the state of knowledge in the following research area:

- the use of Semantic Web technologies to advance characteristics of Access Control, as one of the main approaches to protect data subjects' data from unauthorised access (Section 2.3),
- the specifications of Transparency Enhancement Technologies provide data subjects with more visibility about their data collection, access and processing (Section 2.4).

The integrative review approach [46] is selected to perform in this research because:

- On the one hand, access control is a mature, well-established and well-studied topic with enormous application in different domains. Therefore, it is not suitable and even feasible to review all published literature and look at how research on access control has progressed over time or how a topic has developed across research traditions is irrational and unattainable.
- On the other hand, transparency enhancement has been attracted lots of interest from researchers of different disciplines recently due to several regulatory and business needs. Therefore, a literature review needs to create initial or preliminary conceptualisations to combine perspectives and insights from various research domains.

The next section describes the undertaken methodology for the literature review in this research.

2.2 Literature Review Methodology

There are two common processes to conduct literature reviews: search and selecting articles, and data analysis and synthesis [47]. The integrative review is not mainly developed according to a specific standard [48]; therefore, it often requires a more creative data collection [49]. In this research, we performed a slightly modified approach for the initial search and retrieval of the relevant publication. Instead of searching databases for related papers based on search keyword, recently published (2015- 2018) peer-reviewed survey papers on the above topics were searched. This approach allowed us to save more time and effort according to relying on the preselected set of relevant publications at the beginning of the review process. A further search based on citation analysis (known as backward/forward snowballing) was performed [47]. The high priority of analysis in backward snowballing was given to high-occurrence items in the reference of several preselected surveys. Forward snowballing was performed based on the recommendation of [50]. Beside relativeness, the following criteria applied for screening the derived papers from snowballing:

- The availability of a full-text version of the publication to evaluate its contents was considered a prerequisite.
- Publication type and the reputation of the publisher, journal, or conference were not taken into account while screening the papers.
- The disciplines of the researchers and their affiliations were not considered as a criterion during the screening process.
- Only publications in English were considered.
- Publications were not screened based on the number of pages or their word count, so long as they fulfilled the other selection criteria.
- In order to be considered for inclusion, a publication must be peer-reviewed, ensuring the level of quality and rigour expected from scientific publications.

Qualitative and thematic content analysis was then applied to identify, analyse, and report patterns in the form of themes within the final selection of publication [51].

The remaining of this chapter describes the review of the literature and its findings.

2.3 Access Control

In order to address the highly publicised privacy concerns, Privacy-Enhancing Technologies (PET) have been developed to govern how personal data can be accessed and raise the awareness of data subjects when it comes to sharing their sensitive data. Most data privacy studies [52] [53] categorised Access Control as a PET due to the support they provide to data subjects for construct barriers preventing unauthorised audiences from access to their personal data. In general, access control is used to refer to a framework, which is a combination of

- an access control model, which is a scheme used to guide the access control process;

- the policy language, which defines both the syntax and the semantics of the access control rules; and
- the enforcement mechanism, which deploys the access control rules in the access control process.

The Semantic Web community have had a strong influence on access control research [54], which can mainly be categorised and listed as follows:

- representing existing access control models and standards using semantic technologies;
- proposing new access control models suitable for open, heterogeneous and distributed environments; and
- recommending languages and frameworks that can be used to facilitate access control specification and maintenance.

This section presents relevant access control models and discusses how they were proposed or enhanced using semantic technologies. To keep it consistent with rest of the thesis, "user", "resource" and "resource owner" denote entities requesting access to the data, personal data and data subject, respectively.

2.3.1 Access Control Models

2.3.1.1 Mandatory Access Control (MAC)

MAC, originally developed for military applications, limits access to resources using access control policies determined by a central authority [55]. Users and resources need to be classified based on their security levels by the central authority. Labels are assigned to the resources to represent the security level required to access them. Access is granted to users with the same security level or higher. Therefore, it is best suited to closed environments, where a great deal of control is required [56]. Considering the open, distributed and heterogeneous nature of the web, MAC has not gained much traction among the Semantic Web community, not surprisingly.

2.3.1.2 Discretionary Access Control (DAC)

DAC is generally considered an identity-based access control model where access rights on one or more resources are assigned to users based on their identity. In this model, users have complete control over their access rights on the assigned resources. They can privilege other users the access right on their assigned resources (formally known as a delegation) [57]. DAC is very flexible in assigning access rights between users and resources. However, it is not easy to maintain, follow and control access to the resources and verify the security principles because users can manage their access rights to their owned resources.

2.3.1.3 Role-Based Access Control (RBAC)

RBAC is considered as an alternative approach to MAC and DAC. It restricts access to resources to groups of users with common responsibilities or tasks (roles). In RBAC, users are assigned to appropriate roles and access to resources is granted to one or more roles. Role deactivation is generally used to refer to the process where a user is removed from a role [58].

Depending on the use case, roles may be organised to form either a hierarchy or a partial order [59]. Such structures are used to simplify access control specification and maintenance [60]. Access control constraints are commonly used to enforce conditions over access control policies [61], which can be listed as follows:

- static and dynamic separation of duty (a user cannot be assigned to two roles simultaneously); and
- least privilege (a user can only be assigned to a role if s/he has already been assigned to another required role) [62].

Besides all of these advantages, RBAC also has some drawbacks. It is frequently criticised for the difficulty of setting up an initial role structure, especially in large systems, where role inheritance and the need for customised privileges make administration potentially massive [63]. Another drawback is inflexibility in rapidly changing IT technologies where RBAC provides insufficient support for dynamic attributes like time of day, which might be needed when determining user permission [64]. Most of the research effort within the Semantic Web community focuses on modelling RBAC using ontologies [65] [66] [67]. The main difference between these studies comes from the way that RBAC concepts (User, Role, Permission), their relation, such as role deactivation, and the constraints (separation of duty and least privilege) are modelled.

2.3.1.4 Attribute-Based Access Control (ABAC)

Unlike more traditional access control models, ABAC allows for creating access policies based on the existing attributes of users and resources in the system, rather than the manual assignment of roles, ownership or security labels by a system administrator. An accepted high-level description of ABAC functions is defined in [68] as “an access control method where subject requests to perform operations on objects are granted or denied based on assigned attributes of the subject, assigned attributes of the object, environmental conditions, and a set of policies that are specified in terms of those attributes and conditions”.

Several approaches have attempted to use Semantic Web technologies in the ABAC model. A study proposed a hybrid RBAC-ABAC model with a supporting framework based on a variant of Web Ontology Language (OWL), namely OWL Description Logic (OWL-DL), where attributes are used to classify users into access control roles [69]. While all essential RBAC elements were formalised into an OWL-DL ontology and details for expanding OWL-DL expressiveness with SPARQL[3] were given, the authors did not fully model the attribute-based aspects of their ontology. Details on how attributes are defined, assigned, related to users or how they may be combined with their framework were not provided. Another research described a “Constraint and Attribute-Based Security Framework for Dynamic Role Assignment” focused partly on using a user’s physical location for role assignment [70] [71]. In this approach, predefined roles could have both previously known sets of users and users who dynamically assigned according to the content of their attributes and policy set on role assignment. Rather than employing a policy language like most ABAC works, constraints were defined as attribute-value pairs assigned directly to roles. Semantics for role inheritance and constraint dominance was given in addition to a description of an OWL-DL ontology-based prototype. A Semantic-aware Attribute-Based Access Control model (SABAC) was proposed in [72], which represents attributes by ontology and handles authorisation decision with ontology inference, and XACML (eXtensible Access Control Markup Language) is used to describe the access control policies. Finally, in [73], a Semantic Web-based RBAC model was proposed to add ABAC elements. The model represents and provides a means to reason on hierarchical

[3] https://www.w3.org/TR/sparql11-query/

RBAC in description logic using SWRL and uses attribute-based policies for role assignment.

In general, ABAC is better than RBAC in terms of scalability (i.e. number of users to be managed), flexibility (i.e. it is easier to implement in a large-scale environment) and access control management (i.e. it is easier to associate attributes to other users or resources) [74]. ABAC is created to be a more dynamic and fine-grained model to serve the current large scales systems better. Also, defining access policies among different organisations becomes possible in ABAC due to its attribute-centric nature. Nevertheless, to fully cover all the possible access circumstances, the number of access policies could be enormous, which leads to policy administration problems, such as policy redundancy and policy conflicts [75]. Therefore, a concise but comprehensive set of policies is necessary to save time and efforts for the system and system administrators. In the real-world, policy redundancy and conflicts are common, especially in large-scale systems. As the number of policies increases, manually eliminating incorrect policies and clarifying the semantic meaning of inferred policies is time-consuming and almost impossible to accomplish [76]. So, policy administration plays a vital role in ABAC because the whole access control system relies on policies to protect the integrity, availability, and confidentiality of resources in the system [77]. Recently, a study [78] proposed a new model of ABAC with ontology, OABACM, which is convenient for policy representation and reasoning. An ontology was applied in OABACM to depict entities and their relations in the access control domain intuitively. Different kinds of relations between ABAC entities, including equivalence, inclusion, and disjoints, were identified and described. Inherent logical properties of the model were formalised to improve the efficiency of access policy administration by reducing policy redundancy and detecting policy conflicts.

2.3.1.5 Context-Based Access Control (CBAC)

The need for shifting a paradigm from user-centric to context-centric access control has been recognised in some initial researches [79] [80] [81] [82]. These CBAC approaches usually consider the different types of contextual conditions for making access control decisions which can be grouped as follows:

- Actor-Centric Contexts; are the information about representing actors. An actor can be the user, the resource owner or any other environmental person.

- Resource-Centric Context; are the information about representing data or information resources.
- Environment-Centric Context; are the information about representing the surrounding environment between actor and resource, such as the location from where the access request has been originated.

One of the earliest CBAC models proposed by [83], called UbiCOSM, which can be considered as RBAC-extended model, uses context information to define and enforce access control policies. In the proposed modelling, user identities and roles were specified using logical properties. Policies are expressed the dynamic contextual conditions in order to specify the user-role and role-permission assignment policies. These policies assign the many-to-many mapping between a set of users and roles or between a set of roles and permissions, respectively, when a set of dynamic contextual conditions are satisfied.

In some studies, the contextual conditions were defined by the system administrator through a semantic representation in one of the following methods:

- To define the context in policy ontology in which an entity operates in a specific context automatically acquires the ability to perform the set of actions permitted in the current context [84]. This approach deployed OWL to specify ontologies and SWRL to encode rules.
- To provide a context ontology and a policy ontology to specify positive and negative authorisations and obligations [85]. It is discussed that access was enforced by representing access requests as SPARQL queries executed over the knowledge base. However, it is not clear how changes to contextual information were handled in the proposed approach.

A new context-aware access control to support software services was introduced in [86]. The authors defined a context model to systematically represent and capture different types of context information and a policy model to define and enforce access control policies based on relevant contexts from the context model. Both models are specified by the ontology language OWL and extended with SWRL for inferring implicit context and policies with user-defined rules.

2.3.2 Policy Languages and Enforcement Mechanisms

An access control policy consists of authorisation rules that regulate access to data and resources. At the decision-making time, a request to access a resource is evaluated against the rules in the policy. Access control policies are represented in various policy specification languages. Policy languages can be categorised as general, where the syntax caters for a diverse range of functional requirements (access control, query answering, service discovery, negotiation, to name but a few), or specific, which focuses on just one functional requirement [54]. Two popular choices for specifying policy languages are XML (eXtensible Markup Language) and ontologies. This popularity is due to their flexibility, extensibility and run-time adaptability.

The semantics in XML-based approaches are mostly implicit, which cause ambiguity, promote fragmentation into incompatible representation variations [87].

In contrast, ontologies are better suited to modelling the semantic relationships between entities [87]. Furthermore, the common framework and vocabulary used by ontologies provide greater interpretability and interoperability [88]. A broader comparison of policy languages can be investigated in [89], which evaluates different policy languages against a set of criteria for ensuring security and privacy in a Semantic Web context.

In the rest of this section, policy languages specified by ontologies are investigated in detail and categorised based on the method they apply to present policies in the ontology and their enforcement mechanism.

2.3.2.1 Defining policies through ontologies

Ontologies facilitate the merging of the access control policies represented in different vocabularies and their adoption by others. Also, an ontology-based approach can perform:

- deductive reasoning (deriving the consequent) to infer new policies based on the relationship between access control entities; and
- abductive reasoning (affirming the consequent) to specify the access rights required to match a given policy

over access control policies through standard description logic reasoners.

KAoS (Knowledgeable Agent-oriented System) is one of the most popular general policy languages that adopts a purely ontological approach [90]. It was initially designed to enable interoperability between complex web agents then is used as an open distributed architecture for the specification, management and enforcement of various policies. In initial versions of the language, policies were represented using DARPA Agent Markup Language (DAML) [91]. However, the authors later moved to OWL [92] [88].

It uses ontology concepts (encoded in OWL) to build policies. First, KAoS Policy Ontology (KPO) is loaded. KPO defines a set of core vocabularies that are used to describe *actors* (both humans and artificial agents); *actions* (various system operations such as accessing, communication and monitoring); *resources* (entities associated with actions); *policy-types* (authorisations and obligations); and *policies* (positive and negative constraints). An additional ontology is then loaded on top of it, extending concepts from the generic ontology, with notions specific to the particular controlled environment. The KAoS Policy Service distinguishes between *authorisations* (i.e. constraints that permit or forbid some action) and *obligations* (i.e. constraints that require some action to be performed when a state or event-based trigger occurs or else serve to waive such a requirement). In KAoS, context conditions constraining a policy may be specified by defining appropriate classes defined via property restrictions. The use of OWL enables reasoning about the controlled environment, policy relations and disclosure, policy conflict detection, and harmonisation, as well as about domain structure and concepts exploiting the description logic subsumption and instance classification algorithms. In [93], the authors discuss how description logic can support policy administration, exploration and disclosure. The administration is mainly concerned with subsumption based reasoning and the determination of disjoints. Exploration and disclosure are supported using instance classification capability. Constraints' test and return relevant constraints given one or more properties are possible using abductive reasoning. Nevertheless, a pure OWL approach encounters some difficulties in defining some kinds of policies, especially

- those that need to define constraints over a property with statically unknown values and
- those that contain parametric constraints, which are assigned by a value only at deployment or run-time.

To deal with these issues, KAoS developers have introduced role-value maps as OWL extensions and implementing them within the Java Theorem Prover, used by KAoS [92]. The adoption of role-value maps and using description logic-based concept constructors [94] allowed KAoS to specify constraints between property values expressed in OWL terms and define policy sets. Policy sets were referred to as a group of policies that share a common definition but can be singularly instantiated with different parameters [95]. Conflicts were identified and resolved at design time using deductive reasoning based on policy priorities and timestamps [93].

2.3.2.2 Defining policies through Rules

Support of access control policies that contain instance dependencies or variables is the main advantages of rule-based approaches. However, access control policies specified using different vocabularies still can be integrated since these approaches also define access control policies over ontology entities.

In this section, one of the main rule-based languages and enforcement frameworks is examined, primarily concerned with the specification and enforcement of policies in ubiquitous environments.

Rei [96] [97] is a Semantic Web policy language and distributed enforcement framework, which permits to specify, analyse and reason about declarative policies defined as norms of behaviour [98]. The first version of Rei was defined entirely in first-order logic with logical specifications for introducing domain knowledge using RDFS (Resource Description Framework Schema) or Prolog rules. The authors later provided an OWL representation for their policy language due to richer semantics of OWL compared to RDFS [99] [100], and they adopted a rule-based enforcement mechanism, in contrast to the description logic enforcement mechanism adopted by KAoS. Policies are developed as contextually constrained deontic concepts, i.e. *permission*, *prohibition*, *obligation* and *dispensation,* to restrict domain actions that an entity can/must perform. *Permission* and *prohibition* in Rei are directly mapped with the *positive* and *negative authorisations* in KAoS; likewise, *obligations* and *dispensations* in Rei are mapped with *positive* and *negative obligations* in KAoS.

In Rei, rules are expressed as OWL properties of the policy. Context conditions can be defined as one or boolean combination of a pair of simple constraints. A constraint is associated with a policy at three different levels as follows:

- The first possibility is to impose a constraint within the definition of a deontic entity as a property. In this case, the constraint can be expressed over
  - the actor,

    - the action to be controlled, or
    - generic environmental states, e.g., the time of the day.
- Also, constraints can be imposed within the *Granting* specification over
    - the entity the granting is made to,
    - the deontic entity the granting is made over, or
    - the generic environmental states.
- Finally, it is possible to directly express a set of constraints within the policy definition, which is generically defined as conditions over attributes of entities in the policy domain.

Although represented in OWL-Lite, Rei still allows the definition of constraints follows the typical pattern of rule-based programming languages, i.e. defining a variable and the required value of that variable for the constraint to be satisfied. In this way, Rei overcomes one of the significant limitations of the OWL language, and more generally of description logics, i.e. the inability to define variables. Therefore, Rei's rule-based approach enables defining policies that refer to a dynamically determined value, thus providing greater expressiveness and flexibility to policy specification. Another fundamental property of Rei is its non-monotonic inference due to negation-as-failure. For example, open policies prescribe that authorisations by default are granted, whereas closed policies prescribe that they should be denied unless stated otherwise. Other non-monotonic inferences, such as authorisation inheritance and overriding, are supported as they were in ontology-based policy languages [89].

On the other hand, the rule-based approach of Rei treats Rei rules knowledge separately from OWL ontology knowledge due to their different syntactical form. OWL inference is essentially considered an oracle, i.e. Rei rules cannot be exploited in the reasoning process that infers new conclusions from the OWL existing ontologies. In other words, the Rei engine can reason about domain-specific knowledge but not about policy specification. As a primary consequence of this limitation, Rei policy statements cannot be classified using ontological reasoning. Therefore, in order to classify policies, the variables in the rules need to be instantiated.

Unlike KAoS, Rei cannot statically detect conflicts, but it can only discover them in a particular situation. Given Rei allows for policies to contain variables, conflicts need to be resolved at run-time instead of design time, which is the case with KAoS. In [97], the authors discuss how conflict resolution can be achieved using meta-policies and a partial order between them.

2.3.2.3 Defining policies through Combined Ontology and Rule-Based Approaches

A hybrid approach to policy specification and enforcement can be used to exploit the out of the box deductive capabilities of an ontology-based approach and the run-time inference capabilities of a rule-based approach. On one side, using ontology-based approach exploits description logic to describe contexts and related policies at a high level of abstraction allows their classification and comparison. This feature is essential to detect conflicts between policies before they are enforced and is granting interoperability among entities belonging to different domains that adopt different policies. Another interesting application of an ontology-based approach lies in the possibility of exploiting policy description to facilitate negotiation in policy disclosure. On the other side, a rule-based approach relies on the features of logic programming languages to express contexts and related policies in a clear, concise and expressive way to enable evaluation and reasoning. This section describes Proteus [101], which uses a combined approach to policy enforcement.

Proteus [101] uses a hybrid approach to semantic policy specification. This context-aware adaptive policy model uses ontologies to model both domain information and policies like KAoS, which allows for conflict resolution and harmonisation at design time. It also adopts a rule-based approach to support dynamic constraints and run time variables like Rei.

The Proteus context and policy model are described as the interactions occurring in a system using the concepts of entities and actions. Any *actor* or *resource* in the system is represented as an entity and is logically characterised by several properties expressed as attribute-value pairs. An *action* describes an activity performed by an actor or another entity within a specific operating situation, called the action context. The action context consists of attributes that qualify the action and the entity that is performing then action. An interaction defines an association between an entity and an action.

Proteus models an activating context as a set of attributes and predetermined values labelled in some meaningful way and associated with desirable semantics. An attribute could define constraints either a single value or for a range of allowed values and can be assigned to a fixed constant or a variable over a value domain. An activating context can be minimal, i.e. formed by a single attribute/value pair, or composed of minimal contexts, i.e. defined by multiple attribute/value pairs [84].

In order to deal with possible conflicting situations, Proteus allows the definition of constraints over different activating contexts and distinguishes overlapping and disjoint activating contexts.

Description logic deduction is used to determine the policies that are relevant for the instance data supplied. However, since description logic reasoning is insufficient to cater to contextual properties based on property paths or associated with variables [102], the authors combined it with LP-based reasoning and propose context aggregation and context instantiation rules following the approach described in [103].

In [104], the authors provide details of Proteus prototype implemented in Java with a Pellet reasoner. The proposed solution supports incremental reasoning via an OWL application programming interface and SPARQL queries.

2.3.3 Discussion

This section discussed different access control models and approaches that have been applied in different environments using Semantic Web technologies. As described, one of the substantial limitations of the traditional access control models, such as MAC and DAC, is that they are identity-based access control models which evaluate access decisions based on individuals. This limitation causes a high complexity of security administration and a high cost of managing large-scale systems. Especially, the number of authorisation policies can become extremely large whenever the numbers of users and resources are high. The RBAC models have solved this problem by making access decisions based on groups or roles of individuals. However, these models do not consider the dynamic attributes (i.e. context information), such as the location and request time, and permission assignments to users are static yet.

The ABAC facilitates modelling a wide range of dynamic attributes of access control using the rule-based approach. These models are easy to set up but complex to manage in large-scale systems because of the enormous numbers of attributes. Also, satisfying the dynamically changing contextual conditions in these access control mechanisms are limited to attributes of relevant entities and users' roles. CBAC models consider the combinations of richer context information such as the user, the resource, and their environment-specific conditions to grant access to resources. Most CBAC models have adequate functionalities to incorporate diverse context and situation information into user-role and role-permission assignments for dynamic access control decision making.

The change of contextual conditions such as the temporal, spatial and interpersonal relationship information plays a significant role in maintaining the privacy requirements of the associated stakeholders in today's dynamic era of the digital world. Also, there is always an association between a system (a system can constitute users and resources) and its environment. The system should continuously adapt to the ever-changing dynamic situation of its environment.

Therefore, CBAC seems more appropriate than other approaches to satisfy the need to incorporate a general context into the policy model to provide dynamic access control decisions. Considering all the above advantages of CBAC, it is considered as a preferred access control model for this study.

Another main area of review in this section was focused on technical aspects of the main ontology-based policy languages, the way they specify and enforce the policies and the difference in expressivity, kind of reasoning required, features and implementations provided.

It was argued that the use of ontologies to describe related policies at a high level of abstraction could facilitate their maintenance and adaptation and decreases the number of errors before enforcement. Also, providing important supplemental information through policy description using ontologies can improve negotiation. However, this approach is insufficient to define constraints over a property with statically unknown values or cater subsumption-based reasoning to the instances containing parametric constraints with value assignment at run-time.

On the other hand, the rule-based approach provides greater expressiveness and flexibility to policy specification through dynamically determined value. Another fundamental advantage of rule-based approaches is that their inference is non-monotonic, which allow them to make default decisions in the absence of complete specifications.

However, policy statements in the rule-based approaches cannot be classified using ontological reasoning before instantiation of variables in rules. Therefore, instead of policy conflict detection in design time, they can be only detected and resolved at the run time in rule-based approaches.

The combination of both above approaches has been deployed in literature and discussed in this section. Since the hybrid approach exploits the advantages of both above approaches, and it would be considered in the design of policy specification and enforcement mechanism of this study.

Despite the success of Semantic Web technologies in representing existing access control models and standards, these models still have become ineffective as a privacy protection approach.

Some studies [12] [11] have discussed this deficiency in detail and claimed that the main focus of a privacy protection approach must be on controlling the proper use of the data, which constitutes a whole dimension of privacy concerns for users [105], [106]. Interestingly, as control over data is perceived to be low, it would increase the level of related privacy concerns [107]. Likewise, a perceived higher level of control may increase the willingness to disclose personal information [16].

In [108], the authors applied the theoretical framework of control theory to privacy, considering control of data subjects over information and revealed multiple controllability issues of privacy. This research showed that a data subject would need a sufficient understanding of causal and temporal relationships between their actions and privacy entailing consequences. The authors proposed modelling elements of the privacy decision-making process in more detail as a future research direction.

Data subjects cannot understand the inherent consequences of access control's privacy policy in most cases. They also need to set their preferences for each of their sites/services, which most probably use different terminology [89]. Managing these policies across several systems is impractical. Raising the problem mentioned above, the authors of [109] claimed that "access control in itself is inherently inadequate as a framework for addressing privacy on the Internet". They discussed that information systems/services need to have the following characteristics to support control of data subjects over the responsible use of their private data:

- They should give the data subjects due notice in collecting their data and using it and allowing them to respond appropriately to either take action to preserve their privacy or give it up voluntarily in exchange for better service.
- They should provide a mechanism to track the data as it follows through the system and maintain detailed provenance information in a machine-understandable format to reason over them to identify misuse of data ("post-facto accountability"). This mechanism should include policy tools that not only identify violations but also support daily operations by answering questions about the use of data.

- They should provide all data subjects with accessible and understandable views of the policies associated with information resources ("Policy-awareness") and help them to understand the privacy implications of their actions on the web ("Privacy implications").

The concept usually used to describe these properties is *transparency*, which also is specified as a property of "visibility" in [30]. Therefore, transparency can be considered as a state to mitigate any obstacles, impeding the visibility of the previously disclosed data by a data subject.

2.4 Privacy-relevant data processing Transparency Enhancement

Considering the distributed nature of current online services, harden it to users of such services to keep track of where information about them is stored, to whom it is handed out and for what purposes it is used. Emerging technologies such as cloud computing, IoT and AI intensify this situation. As argued in the previous section, the traditional approaches need to enhance to focus on controlling the proper use of the data. To this end, a data subject must get information on how and why her personal data is used and possibly from which sources it originated. Several EU projects (FIDIS[4], PRIME[5]) were proposed and conducted to construct concepts and tools that can help data subjects to regain control over their private sphere and to understand the consequences to their privacy in an online networked world. The concept of Transparency Enhancing Technologies (TETs) for privacy purposes was initially defined in one of the deliverables of the FIDIS project [110]. However, this definition was too narrow and provisional, considering the implications of the word transparency. In [111], TETs were defined as tools that can provide concerned data subjects with clear visibility on aspects relevant to their personal data and privacy. This definition was leveraged in [112] [113] to define TETs as technological tools that provide data subjects, or a proxy that acts on behalf of them, with information on (intended) data collection, storage, processing and disclosure in an accurate and comprehensible way.

The rest of this section will discuss how transparency tools enhance data subjects' control over gathering and processing their personal data through an integrative literature review. This review leads to identify a set of categorisation parameters for

[4] FIDIS (Future of Identity in the Information Society), http://www.fidis.net

[5] Prime Life, http://www.primelife.eu/

describing the properties and functionality of proposed approaches. Different purposes of enhancing transparency and the corresponding type of representation per each purpose are investigated in detail. Also, various measures to evaluate the impact of the proposed transparency enhancement approach are reviewed and discussed.

2.4.1 Categorisation of Transparency-Enhancing Technologies

The need for common terminology to precisely describe the properties, requirements and functionality of TETs has been identified and addressed in several studies [112] [113] [114] [115]. This common terminology can facilitate the comparison of proposed TETs with currently existing ones. In this section, the development procedure of existing terminologies is discussed briefly, and their advantages and disadvantages are reviewed. They are then synthesised to analyse TET's suitability to contribute to data subjects' privacy preservation.

The first categorisation of TETs is provided in [114] as:

- **Type A:** "legal and technological instruments that provide (a right of) access to data processing, implying a transfer of knowledge from data controller to data subject, and /or"
- **Type B:** "legal and technological instruments that (provide a right to) counter-profile the smart environment to 'guess' how one's data match relevant group profiles that may affect one's risk and opportunities, implying that the observable and machine readable behaviour of one's environment provides enough information to anticipate the implications of one's behaviour."

Although this categorisation of TETs has the advantage of including both technological instruments and legal approaches towards transparency, it does not provide further insight into TET's technical functionality and properties. It is too coarse-grained to assess and analyse the suitability of technological TETs.

Another approach of TETs categorisation [112] proposed different classification parameters focusing more on the technological aspects of TETs. Although the research provides a broad description of TETs, it does not clearly define some parameters. The lack of clarity in parameter definition entails imprecise classification of TETs and makes it difficult to compare different TETs relying on this classification.

In [113], an overview of a selection of existing TETs was provided along with a high-level categorisation of TETs. The presented TET categories were specified based on the transparent insights they provide as:

- Transparency as insight in intended data collection, storage, processing and/or disclosure
- Transparency as insight in collected and/or stored data
- Transparency as insight in third party tracking (insight in user behaviour data disclosure)
- Transparency as insight in data collection, storage, processing and/or disclosure based on the website's reputation
- Transparency as insight in (possibly) unwanted user's data disclosure (awareness promoting)

While this categorisation specifies some information about the type of transparency provided by TETs, no details are provided about deployed technology and the type of data to support the transparency.

Later, in [115], the authors discussed and compared all approaches mentioned earlier. They collected and presented a set of categorisation parameters for describing the properties and functionality TETs as follows:

- **Application Time (AT):** This parameter defines the time in which transparency can be provided regarding data collection and processing.
    - **Ex-ante transparency** informs the intended data collection and processing, thus enabling the anticipation of consequences before disclosing data.
    - **Ex-post transparency** offers insight into what data was collected, processed by who, disclosed to whom, and whether the data processing has been in conformance with negotiated or stated policies and can be informed about consequences.
- **Target Audience (TA):** The expected users of TETs can be divided based on their expertise (professional and non-professional) and categorised as
    - the **auditors (data controller)**, people/proxies that do professional audits for privacy protection
    - the **data subjects**, whose personal data is collected and processed.

TETs for data subjects are expected to have a high level of "user-friendliness" with the presentation of information in a manner that is easy to understand. The privacy implications of different choices and actions should be explained so that data subjects can understand their actions and related consequences. Therefore, these tools usually try to find alternative ways of presenting complex properties with limited information. Tools targeted towards auditors/proxies produce their logs and audit trails to understand what data decisions are made.

- **Environmental Context (EC):** This parameter specifies a differentiation based on the environment the TETs are designed for and defines technological and functional constraints of TETs' design.
    - In a **solitary environment**, the TET functionality focuses on reviewing the data subject's personal data independently of other users of the same service.
    - In **participatory communities**, the TETs concentrate on certain facets of data subjects' lives exchanged with other participants of the same social circle.
- **Interactivity Level (IL):** This parameter is adopted from [116] and extended "Possibilities of Control and Verification" parameter of [112] from a technological perspective to describe how a data subject or an auditor interacts with a TET (**Passive Read-Only, Interactive Read-Only and Interactive**). In interactive TETs, the level of control that s/he can apply through a TET (**Collection, Usage, Modification and Deletion**) can be specified.
- **Delivery Mode (DM):** This parameter describes the notification method of data subjects/auditors about relevant aspects of their privacy. TETs can either actively notify data-subjects (**Push**) of events relevant to their privacy or wait for data subjects/auditors to ask for notifying them (**Pull**) actively.
- **Data Types Presented (DT):** This parameter defines what type of data can be gathered and used by TETs. The parameter's manifestations are adopted from [117] [118] [116] and can be defined as follows:
    - **Volunteered Data**; "data which a user actively and knowingly discloses",
    - **Observed Data**; "data a user passively discloses, i.e. data that results from the interaction of a user with a provider.",

    - **Incidental Data:** “data about a user that is disclosed not by the user herself but by others.”
    - **Derived Data**: “data about users that is inferred as a result of data analysis.”

- **Assurance Level (AL):** This parameter describes the extent to which data subjects can determine the completeness and correctness of the information provided by a TET. Unlike **untrusted TETs**, the correctness and completeness of the information provided by **trusted TETs** can be guaranteed by data subjects or an auditing entity using technical means. **Semi-Trusted TETs** provide information in a way that their correctness and completeness cannot be guaranteed by technical means. However, a data-subject or an auditor can manually determine whether the information is correct and complete.

2.4.2 Presentation of Transparency

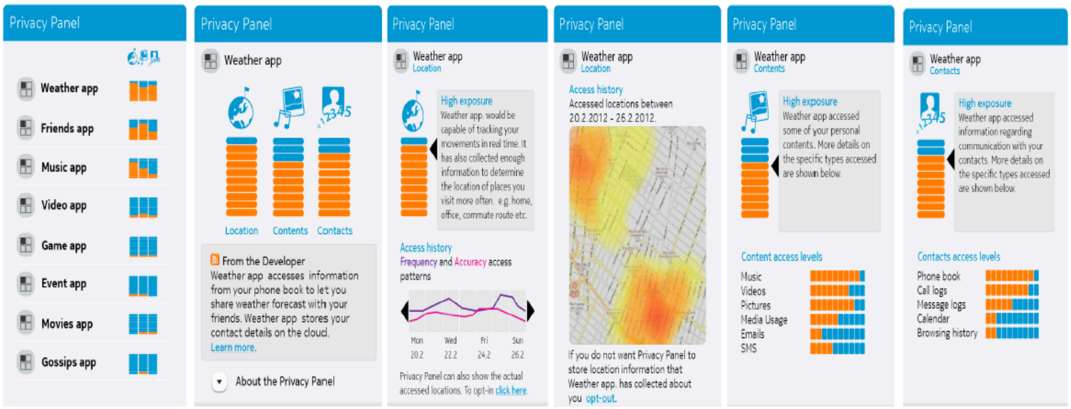
Different TETs in reviewed literature have deployed various representation of transparency for privacy purpose. Most of the researches on TETs have composed textual information to a certain extent with their data visualisation. While some of the studies used sketches or mock-up screenshots to provide transparency, another group of studies reported their use cases and implementation of their platform using graphical elements. This section discusses the different purposes of transparency representation with regards to their application and specification.

2.4.2.1 Expose

Different forms of on-screen representation used by the reviewed TETs to convey meaningful information about the process of recognising disclosed personal data vary widely. The diversity of the various approaches stems most likely from the different usage contexts and the authors’ design preferences during the planning and implementing of their respective prototypes. While relatively few TETs rely solely on textual information, most TETs combine text with graphical information to visualise disclosed personal data or metadata. Colour codes are frequently used to emphasise the meaning of text and graphics [119] [120], [121], [122], [123], [124]. Several TETs use bar graphs, either coloured or monochrome, to signify meaning associated with the properties of disclosed personal data [125], [126]. While a study presents quantitative values via segmented bars instead of continuous bars [127] (Figure 2.1), another one correlates the scalar values of data items to the radius of circles in a bubble chart [128].

To signifying interrelations between multiple stakeholders in participatory communities, some studies used connect stylised nodes [129] [130], while others represented them by bundles of lines between the communication endpoints. Another study [131] used a directed line graph to signify a hierarchical structure of dependencies (Figure 2.2).

In many cases, a specific form of graphical representation is required to visualise the underlying functionality of TETs depending on the particular context of usage. For example, most of the researches designed the TET to enhance the transparency of location-based services, overlay the standard map views of established web-based services, such as Google Maps [126], [124], [128] or OpenStreetMap [133], with additional contextually enriched information. Figure 2.3 shows how a TET deploys this approach to signify blurry areas where a particular person can be found without revealing his/her exact location.



Weather app accesses information from your phone book to let you share weather forecast with your friends. Weather app stores your contact details on the cloud. Learn more.

Weather app. would be capable of tracking your movements in real time. It has also collected enough information to determine the location of places you visit more often. e.g. home, office, commute route etc.

Frequency and Accuracy access patterns

Privacy Panel can also show the actual accessed locations. To opt-in click here.

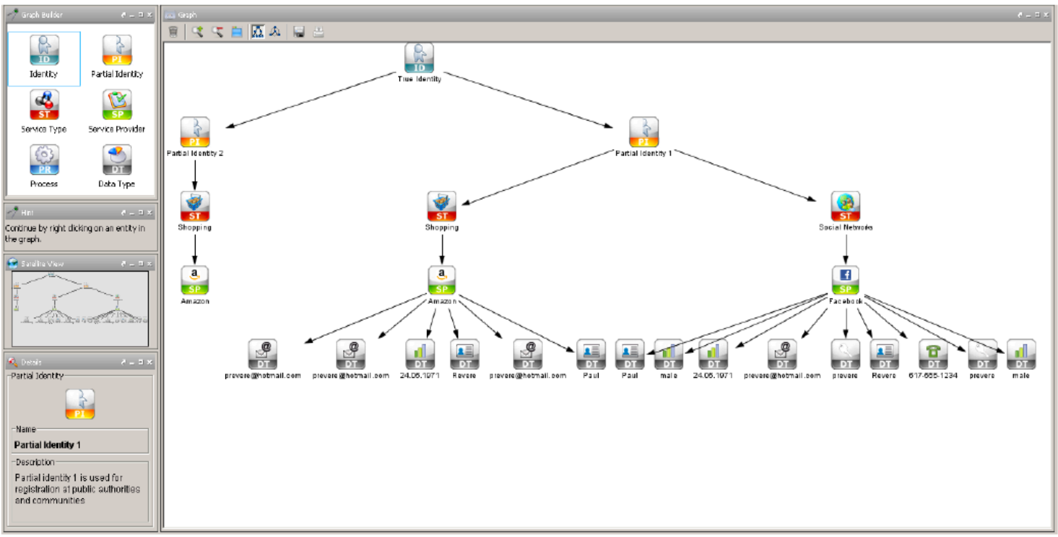
Accessed locations between 20.2.2012 - 26.2.2012.

If you do not want Privacy Panel to store location information that Weather app. has collected about you opt-out.

Weather app accessed some of your personal contents. More details on the specific types accessed are shown below.

Weather app accessed information regarding communication with your contacts. More details on the specific types accessed are shown below.

Figure 2.1. A sample presentation of the properties of disclosed personal data through segmented bars [117]



Continue by right clicking on an entity in the graph.

Partial Identity

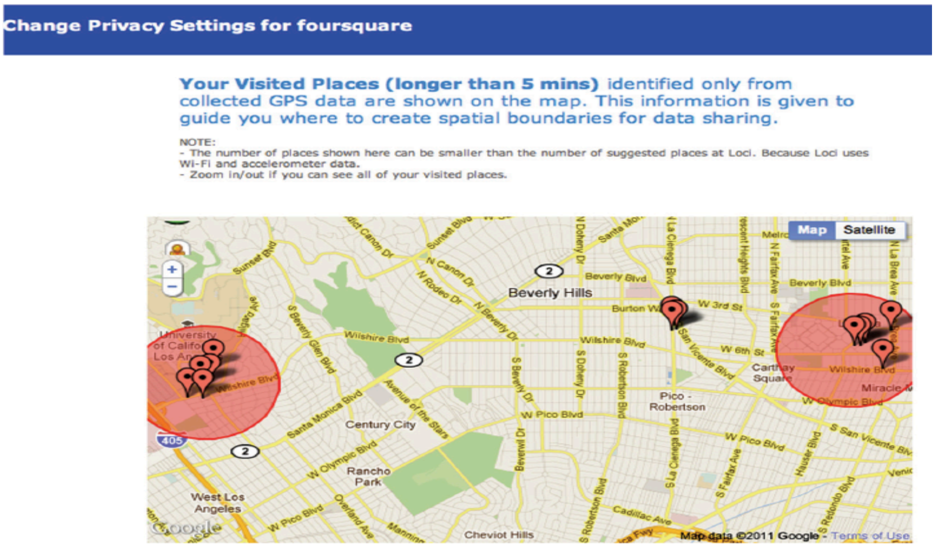
Name

Partial Identity 1

Description

Partial identity 1 is used for registration at public authorities and communities

Figure 2.2. A sample presentation of the properties of disclosed personal data through directed line graph and icons [121]



Change Privacy Settings for foursquare

Your Visited Places (longer than 5 mins) identified only from collected GPS data are shown on the map. This information is given to guide you where to create spatial boundaries for data sharing.

NOTE:
- The number of places shown here can be smaller than the number of suggested places at Loci. Because Loci uses Wi-Fi and accelerometer data.
- Zoom in/out if you can see all of your visited places.

Figure 2.3. A specific form of graphical representation for enhancing the transparency of location-based services [124]

As shown in Figure 2.2, another approach can use icons to complement the textual or graphical visualisations of TETs [131] [125]. According to recognisability of icons, based on previous knowledge of the user or as a result of repetitive exposure of that icon in the same application context, they were used to represent nodes in hierarchical structures, to hint at the underlying functionality of editing and modifying contents, and to denote multiple entity types.

2.4.2.2 Guidance and Awareness

Sometimes TETs go beyond presenting details about disclosed personal data and may guide the user towards better awareness of a particular circumstance or even nudge her to take some critical action about specific disclosed data and encourages her to reconsider her previous decisions. For example, the TET presented in [123] changes their display's colour, if the total amount of disclosures of a trait exceeds a certain threshold. Likewise, another TET changes the taskbar icon's colour to indicate that the user's personal data have been queried [134].

Another group of TETs personalise user's preferences on the severity of disclosing particular data items applying machine learning methods [119] [120] [126] [124]. Some of these methods can predict future data disclosure decisions with high accuracy according to analysing users' decisions in the past or even allowing trained classifiers to make decisions autonomously. Although these approaches facilitate the user's decision-making process by offering favourable options; however, the fundamental functionality of automated decision-making processes may not be transparent to the majority of users. The TET presented in [119] is built based on the concept of "user-controllable policy learning", a cooperative approach between the user and the policy management system. Based on the user's settings, the system makes automated decisions about disclosing user's data according to the request of third parties. Users help to train future decisions of the system by reviewing and commenting on the choices made by the system and applying incremental updates. Similarly, the TET presented in [124] provides users with an active 'rule recommender' that suggests changes to users' privacy settings. The TET is based on an architecture that implements a 'personal data vault' and visualises possible risks based on the user's settings. Likewise, the recommender system of the TET presented in [120] aids users in refining their privacy settings by providing them with meaningful suggestions for changes. In [126], user sharing behaviour is learned through an extensive online survey, and the result was used to measure the accuracy of the automated decision-making of their TET.

2.4.3 User Studies for TETs Evaluation

User studies are conducted during different phases of design or implementation of the presented prototypes of TETs. Pre-design user studies usually are undertaken to understand the previous knowledge, preferences and expectations of the intended target audience. Post-design user studies evaluate the extent to which TETs perform the expected task of enhancing transparency for privacy purposes. This section discusses different measures that were evaluated in user studies for developing TETs across the literature.

2.4.3.1 Usability

Effectiveness, as success in producing desired or intended result [135] and **satisfaction**, as personal comfort, encouragement, and the perceived usefulness of an application, are specified as main measures to assess the usability of interaction systems [136]. Different methods such as observation, questionnaires and interviews are used solely or jointly to assess whether and to what extent an implemented TET is usable by the respective user group. In [128], a qualitative evaluation approach shows that the participants felt that the evaluated TET effectively achieved its intended goal. A gamification approach is conducted to elicit the effectiveness of the notification mechanism and its UI quantitatively in [125]. The authors discuss that the TET informs the participants effectively and meaningfully. They argue that the approach is more comfortable to use than TETs that rely on the retrospective analysis of access logs. In [137], System Usability Scale and a User Experience Questionnaire are used to assess the usability of the proposed TET in comparison with other related approaches. The user study in [126] is conducted through customised questionnaires, some of whose follow-up questions depend on answers given previously. The effectiveness of the policy recommender of the TET in [124] is evaluated by monitoring the number of participants who adapt their settings after being notified by the tool. According to the user study result, the authors conclude that their recommendations effectively help participants establish settings that met their requirements.

The interviews are conducted with participants in [131] to evaluate the usability of the developed TET. The results show that they “understood and valued the advantages” of the tool.

The satisfaction of users is evaluated through the results of different customised questionnaires in several studies. The majority of user study participants in [123] found the TET useful and would like to install it, or a similar app, on a mobile device. Likewise, the user studies participants in [124] and [122] state that the respective TET represented a useful specification for managing their data and were willing to use it in the future. The Authors of [122] report that about one-third of the test subjects used the tool one month after the study had finished. The usability tests conducted in [138] show that most participants found the Data Track as a potentially useful tool, appreciated its transparency options, and would use it regularly.

2.4.3.2 Comprehensibility

Comprehensibility as an ability to be understood is related to some design principles such as "suitability for the task", "self-descriptiveness", and "conformity with the expectation for the user" [139]. Lack of comprehension and consciousness entails mitigating the ability of users to correctly interpret their privacy status and apply sufficient control to change that status as a result of a rational decision [140] [139]. So, the goal of the user studies to evaluate comprehensibility is to distinguish confirmation of the user's mental model with the functionality provided by a TET. However, investigation of literature depicts that only a few studies consider evaluating users' ability to comprehend the process and data follow visualised by the TETs. The majority of these evaluations were done through self-expression questionnaires. In the user studies conducted in [131] and [134], participants stated that they found the user control of both tools "intuitive". Also, the participants of these studies declare that all information was clearly presented [131], and the TET was "easy to use and understand" [134]. Some user study participant in [124] stated that they appreciate the TET functionality but prefer a more intuitive UI. A study conducted in [123] indicates that users had difficulties understanding the technical terms that the TET designers had chosen. Evaluations of 'Data Track' ('GenomSynlig') [138] revealed that test subjects had difficulty differentiating their data access rights and data flow in the client side and server side from the TET's user interface.

2.4.4 Discussion

The outcome of the literature review in TETs for privacy purpose shows a broad level of their maturity across a diverse range of applications. Location-based services could be identified as the most popular target application of TETs during the early years of the last decade due to the growth of mobile technology. However, this trend moved toward most emerging technologies such as cloud-based services [141] [142] and IoT applications [143] [144] [145] lately.

While some studies have developed a prototype for their proposed approaches [146], others have proposed the idea through mock-ups. Their authors reported that the prototypes are under development and implementation of the actual tool is ongoing [147]. In some cases, it remains unclear whether the illustrated visualisations represent actual UIs of implemented prototype or preliminary mock-ups. Also, technical specifications of different functions of the tool were not described clearly in some of the publications [130] [129] [148] [149]. Moreover, no user studies were conducted to evaluate these approaches.

As discussed earlier, the main goal of enhancing transparency in most of the reviewed studies is to allow users to get information about their disclosed personal data and a particular circumstance of their disclosure. In other words, these studies try to develop/propose a tool/method which helps users to be aware of what happens with their personal data, i.e. what data about them are collected and how those data are further processed, by whom, and for what purposes. However, according to the best of our knowledge, none of these studies has provided policy-awareness with users to have an accessible and understandable view of the policies associated with their data. This awareness allows users to know the reason for their personal data disclosure and access and enhance the level of transparency of the corresponding application/service. Lack of transparency on the causality of data disclosure does not seem to achieve "conformity with user expectations" [139].

Finally, as discussed in Section 2.4.3, most of the user studies conducted in reviewed literature measured usability, satisfaction and comprehensibility of TETs based on arbitrary questionnaires for participants' self-expression. Refusing standard approaches/measures to conduct user studies in some of these studies mitigates the validity of their results. Also, not considering participants' mental model and cognition in the user study design led to unreliable evaluation. Further and detailed literature review will be provided in Chapter 5 of this thesis to develop a better view of the specifications of appropriate user studies for assessing relevant measures of transparency.

2.5 Conclusion

This chapter described how conducting an integrative review of pertinent literature fulfils RO1. The contribution of this review is assessing, critique, and synthesise the literature on two topics of interest (access control and transparency enhancement) in a way that enables to map a field of research, reveal the gaps, motivate the aim of the study and justify the research question and hypotheses.

Review of publications on access control manifested the capabilities and high potential of Semantic Web technologies for representing and proposing the access control models and for facilitating access control specification and maintenance. Comparing different approaches showed that the best practice of deploying semantic technologies in the access control model is context-based approaches where the privacy policies are defining through the hybrid approach of using ontologies and rules.

Further, this review highlighted the need to control the proper use of the data and the necessity of giving this control to data subjects to better preserve their privacy. It discussed that this requirement can be fulfilled through providing data subjects with policy awareness and privacy implication which access control in itself is inherently inadequate to provide them. These findings justify RO2 and RO3 of this research and bring up the need to improve the current access control approaches to provide data subjects with more transparent information on how and why their personal data is used. Meanwhile, these findings led us to review the literature on transparency enhancing technologies. This review resulted in the advancement of knowledge about existing approaches of TETs, their categorisation, representation and evaluation. These results featured that most of the existing TETs have been focused on privacy implication to help data subjects be aware of what happens with their personal data. Nevertheless, there is a rare (even not any) approach found to provide policy awareness to enhance the level of transparency to allow data subjects to know the reason for their personal data disclosure and access. These outcomes justify RO2 and RO3 of this research which will be addressed in subsequent chapters.

Also, the review assisted in specifying different metrics used to evaluate TETs; meanwhile, it demonstrated the need for a more comprehensive user-centric evaluation to measure the impact of these technologies. This outcome of review legitimises RO4 of this research which will be addressed in details in Chapter 5.

3 Design

3.1 Introduction

This chapter describes the design and architecture of a proposed novel service, called "eXplainable Personal Data Access" (XPDA), which exploits context-awareness and Semantic Web technologies to provide data subjects with transparency on their personal data access in an interpretable manner. In this sense, XPDA adopts a context-based access control model for the specification of data access policies and deploys a combination of ontology-based and rule-based approaches to enforce and evaluate these policies. Then, it builds the interpretable explanation per each data access decision in an understandable form as a transparency enhancement technology.

Drawing inspiration from the Semantic Context-Based Access Control (SCBAC) model [85], context is considered as a key player for policy adaptation in this service. The term policy adaptation refers to the ability to adjust policy specifications and evaluation to enable their enforcement in different, possibly unforeseen situations, i.e. the context, and to define the expected set of actions permitted based on such context variations [102]. Since the willingness of data subjects to share their personal data and the conditions that their data can be accessed or shared may be mostly unpredictable, policies cannot be specified in advance to cover all run-time situations. Consequently, policies may require a dynamic adaptation to be able to control access to resources.

Another fundamental property of the access control model in XPDA is adopting the combined approach, discussed in Section 2.3.2.3, to policy specification and enforcement. It exploits the deductive capabilities of an ontological approach to enable static policy specification, evaluation and conflict detection. Meanwhile, run-time inference capabilities of a rule-based approach are deployed to evaluate policies based on context variables, whose value is unknown at policy specification time. Therefore, it enables the efficient enforcement of policies defined over dynamically determined context values.

XPDA provides data subjects with ex-post transparency through understandable insights about their personal data access and conformance of the access with the policies. These insights are explained in a manner in which data subjects can understand the detail and cause of a data access decision. This explanation is provided by summarising the current state and their conformity with privacy rules and presented in easy to understand and straightforward presentations to improve the interpretability and usability of XPDA and satisfaction of the data subject.

In the remaining of this chapter, a more detailed decomposition of the architectural building blocks of XPDA would be modelled and documented using the lean graphical notation of the C4 model[6].

3.2 C4 Model

The C4 model is inspired by the Unified Modelling Language (UML)[7] and the 4+1 view model for software architecture [150] to describe and understand how a software system works and to minimise the gap between the software architecture model/description and the source code. It provides a standard set of abstractions as follows to create a ubiquitous language to describe the static structure of a software system. Visualising this hierarchy of abstractions is then done by creating a collection of Context, Container, Component and (optionally) Code (e.g. UML class) diagrams. This is where the C4 model gets its name from:

- **Context diagram** is a starting point for diagramming and documenting a zoomed-out view showing a big picture of the system landscape. The focus should be on the relation between a software system as a black-box and its users or the other systems that it interacts with rather than technologies, protocols and other low-level details.
- **Container diagram** shows the high-level shape of the software architecture, how responsibilities are distributed across it and how the containers communicate with one another. It is a simple diagram to zoom-in on the system boundary focusing on major high-level technology choices.
- **Component diagram** decomposes each container further to identify the major structural building blocks and shows their responsibilities and interactions.
- **Code diagram** is an optional level of detail that can zoom in to each component to show how it is implemented as code, using UML class diagrams, entity-relationship diagrams or similar.

In the following sections of this chapter, the architectural model of XPDA will be discussed through the context, container and component diagrams in further detail.

[6] https://c4model.com/

[7] https://www.uml.org/

Also, inter-communication between containers of the XPDA would be discussed through its sequence diagrams.

3.3 XPDA's Context Diagram

The context diagram displays how the XPDA service interacts with external systems and actors at a very high level. As illustrated in Figure 3.1, the context diagram pictures the XPDA service at the centre as ***XPDA core*** with no details of its interior structure, surrounded by all its interacting systems and actors.

In this diagram, one of the main entities, which XPDA interoperates, is ***Client API***. It is considered as a software system which performs "any operation or set of operations on personal data or sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction" [25].

XPDA core is responsible for controlling this "processing" of personal data. Therefore, XPDA acts as a gatekeeper to control any access and process conducted through ***Client API*** over any data, including personal data stored in ***Internal Data Sources***. All "principles relating to the processing of personal data" [25], including privacy rules, are also in ***Internal Data Sources*** and exploited by ***XPDA Core***.

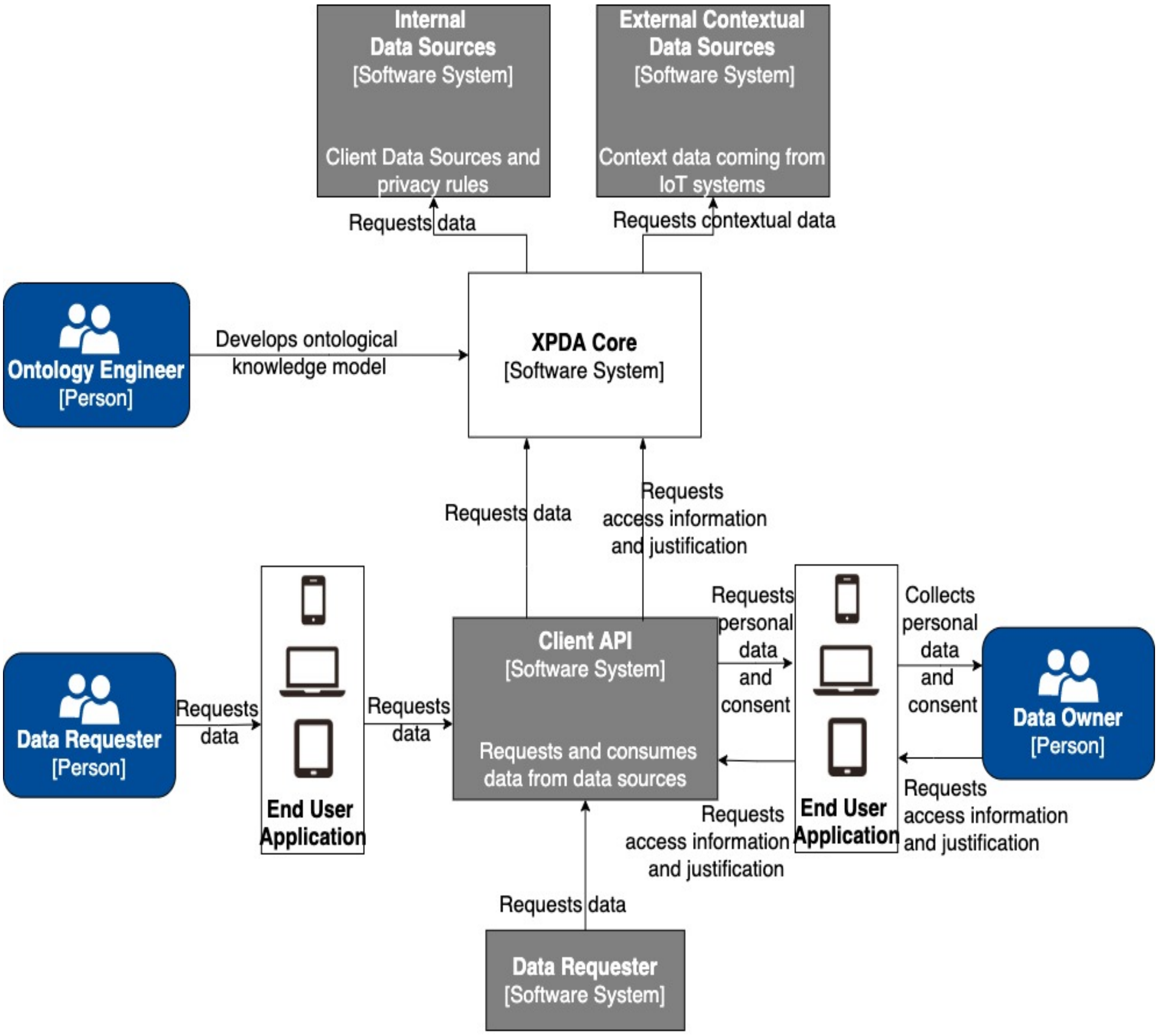


Figure 3.1. Context diagram of XPDA

To apply dynamic policy adaptation in XPDA, external contextual data such as location need to be gathered through various platforms and software systems and transmitted to ***XPDA core***. These systems are demonstrated as ***External Contextual Data Sources*** in the diagram.

Any authorised ***Client API*** user**,** which wants to access or process the data, is considered as ***Data Requester*** in this diagram. The data request can be posted through any ***End User Application*** such as mobile, web or desktop application. Likewise, any request to data disclosure, share or transmission happens for lawful data access and processing by third party software system/service can be posted through ***Client API***.

According to legal regulations, data subjects are playing a pivotal role in any data processing. They need to signify agreement, by a statement or by a clear affirmative action, to the processing of personal data relating to them. This agreement is known as a "consent" and needs to expose a specific, informed and unambiguous indication of their wishes [25]. In XPDA, a data subject, which is illustrated as a ***Data Owner,*** not only interact with ***Client API*** to give his/her consent and disclose his/her data to the system but also is provided with the information relating to his/her personal data access and processing in a concise, transparent, intelligible form, using clear and plain language.

Since XPDA is designed to exploit the capabilities of the Semantic Web technologies, the client application and its underlying data need to be represented in an appropriate machine-readable format. To this end, ***Ontology Engineer,*** as an expert person with experience in Semantic Web technologies and a relevant domain, is responsible for developing an ontological model of the current system and domain knowledge in an appropriate format. S/he is also involved in translating the relevant privacy and business rules of data access identified and specified by regulations or data subjects. S/he needs to provide sufficient specificity and concept coverage to ensure that the ontology is complete, can support its use cases and is current with domain knowledge.

3.4 XPDA's Container Diagram

As illustrated in Figure 3.2, the ***XPDA core*** consists of three fundamental containers, namely ***Knowledge Modelling Unit (KMU)***, ***Access Control Unit (ACU)*** and ***Access Transparency Unit (ATU)***, which can be described as follows:

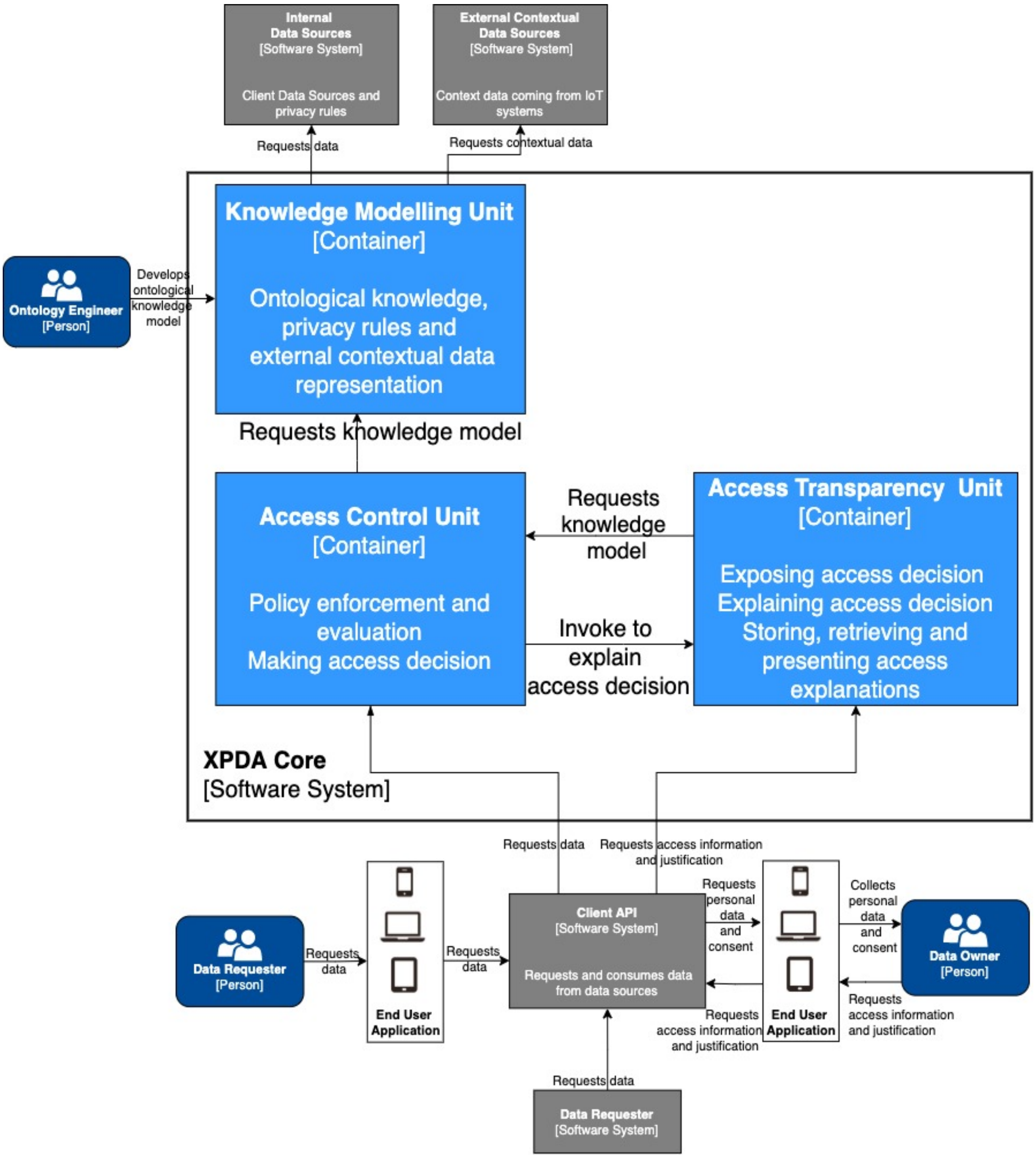

Internal
Data Sources
[Software System]
Client Data Sources and privacy rules
External Contextual
Data Sources
[Software System]
Context data coming from IoT systems
Requests data
Requests contextual data
Ontology Engineer
[Person]
Develops ontological knowledge model
Knowledge Modelling Unit
[Container]
Ontological knowledge, privacy rules and external contextual data representation
Requests knowledge model
Access Control Unit
[Container]
Policy enforcement and evaluation
Making access decision
Requests knowledge model
Invoke to explain access decision
Access Transparency Unit
[Container]
Exposing access decision
Explaining access decision
Storing, retrieving and presenting access explanations
XPDA Core
[Software System]
Requests data
Requests access information and justification
Data Requester
[Person]
Requests data
End User Application
Requests data
Client API
[Software System]
Requests and consumes data from data sources
Requests personal data and consent
Requests access information and justification
End User Application
Collects personal data and consent
Requests access information and justification
Data Owner
[Person]
Requests data
Data Requester
[Software System]

Figure 3.2. Container diagram of XPDA

- ***Knowledge Modelling Unit*** is responsible for making formal conceptualisation and representation of the knowledge within the client application (and its corresponding enterprise or business procedures) and the knowledge provided by external contextual data resources using machine-interpretable ontologies. It concerns the ontology development process, the ontology life cycle, the methods and methodologies of ontology building, and the tools and languages that support them. Beyond domain knowledge representation, another functionality of this unit is to express and represent the privacy policies defined by business, or regulation or data subjects. Policies are usually written in the form of restricted privacy rules in which the permission can be returned as a "deny" or "grant" decision. Privacy rules are specified by representing ontological associations between access decision and contextual knowledge. Therefore, the antecedent of a rule represents the conditions stated in the data specifications, and its consequent represents the entailed access decision.

As the ontological model needs to be expanded, it is performed through merging domain ontologies by hand-tuning each entity or using a combination of software merging and hand-tuning.

- ***Access Control Unit*** specifies the access decisions that allow requesters to access personal data depending on various conditions regarding the contextual aspects. This container adopts a combined approach, discussed in Section 2.3.2.3, to policy enforcement and evaluation, which results in static classification and conflict resolution of context and policy ontologies as well as dynamic evaluation at request time. Privacy rules can be enforced by associating a set of access decisions with specific instantiated contextual conditions in real-time. The rules can be instantiated by adapting the current state to obtain the set of applicable policies. The contexts of applicable policies are verified against the current state of contextual elements to determine the set of currently active policies in the policy evaluation stage. Policy enforcement and evaluation can be triggered by a new request to access the data.
- ***Access Transparency Unit*** provides ***Data Owner*** with more visibility on the required information to justify his/her personal data access decisions, particularly when unexpected decisions are made. This unit not only provides the justifications in order to comply with the "right to explanation" [25], it also warrants there is an auditable and provable way to defend access decisions, which leads to building trust and enabling an enhanced control. Since explanations are forms of social interactions [151], their efficacy and quality mainly depend on their intelligibility and comprehensibility as perceived by data subjects. In other words, an explanation is only useful if data subjects can understand it. As discussed in [152], causal information about a decision that

explains the relation between an input and an output is privileged because it is often favoured and seems to be less cognitively demanding. Providing natural language explanations, which state information about the essential features in a decision, is more relevant than any other method to the causal explanations [153]. The use of non-propositional representations such as diagrams, graphs, and maps present another clear case of the causal explanations [154].

This unit provides a causal explanation for data access decisions and presents them in the above representations.

3.5 XPDA's Component Diagram

Structural building blocks of each container mentioned above, and their responsibilities and interactions are demonstrated in Figure 3.3, which can be described as follows:

3.5.1 Components of Knowledge Modelling Unit

As discussed earlier, ***Knowledge Modelling Unit*** deals with representing the knowledge within the client application by developing the ontologies. This functionality is conducted in ***ontological Knowledge Representer (KR)*** by ***Ontology Engineer***. A domain ontology represents concepts that belong to a realm of the domain. The relevant upper ontologies can also be deployed for modelling the commonly shared relations and objects that are generally applicable across the domain ontologies and for overarching the terms and associated object descriptions in various relevant domain ontologies. The conceptualisation of domain knowledge can be extracted from ***Internal Data Sources*** such as existing databases. Also, privacy rules determined either by regulations, enterprise/application policies or through the explicit consent of data subjects, need to be represented in an appropriate and compatible language in ***KR***.

Meanwhile, some domain knowledge may need to be collected from ***External Contextual Data Sources*** like IoT systems. ***Context Manager (CM)*** collects environmental context information from the corresponding external context acquisition module and formalises them in the same assertion format to aggregate to the knowledge model. ***CM*** can update current state information on event-basis, when any relevant change happens or on demand-basis, where the current state is re-evaluated upon receiving a request.

3.5.2 Components of Access Control Unit

Access Control Unit supports access policy enforcement and evaluation based on current contextual conditions. To fulfil this function, ***Knowledge Manager (KM)*** interacts with ***Knowledge Modelling Unit*** to get the current situation of the ontological model based on updated assertions of the system. ***KM*** is also responsible for updating the model with the inferred knowledge from the ***Reasoning Engine (RE)*** and a new access request from ***Policy Enforcement Manager (PEM)*** and sharing the updated version of knowledge with other XPDA components when it is required. ***RE*** performs reasoning over the ontology to infer new knowledge and determine appropriate privacy rules according to the current state of the context. The reasoning is configured to perform on an on-demand basis in response to incoming access requests. ***KM*** triggers ***RE***, and the inferred knowledge is sent back to ***KM***. ***PEM*** intercepts the access request from ***Client API***, translates it to the same representation of the knowledge base, and dispatches it to ***KM***. ***PEM*** can query the access decision to enforce the privacy rules for evaluating this request with the current context of the system. The result of the access decision is reported to ***Access Transparency Unit*** as well as requesting entity. If access is granted to the request, then the data can be shared with the requesting entity.

3.5.3 Components of Access Transparency Unit

Access Transparency Unit provides ***Data Owner*** with more visibility on details of their personal data access. When a decision is made, ***Access Exposer (AE)*** collects all detail information about the decision from ***PEM***, translates it into a proper presentation, which is more interpretable to the typical person without any specific knowledge of Semantic Web technologies and stores this presentation in ***Log*** database. When a ***Data Owner*** or any controller proxy on behalf of him/her wants to know the details of his/her data access, ***AE*** can retrieve the corresponding presentations of relevant data accesses from ***Log*** and exposes them. ***Access eXplainer (AX)*** is in charge of explaining the reason for a particular access decision according to the corresponding policy enforcement and evaluation. Therefore, an appropriate justification needs to explain which particular contextual circumstances are conformed with a set of defined privacy rules to make this access decision. So, ***AX*** needs to get the current state of the knowledge from ***KM***, provides a causal justification for the access decision, explains this justification in an interpretable representation for novice persons and stores this explanation in ***Log*** for later reference. The explanation can be retrieved from ***Log*** whenever a ***Data Owner*** or any controller proxy on behalf of him/her seeks the reason for personal data access.

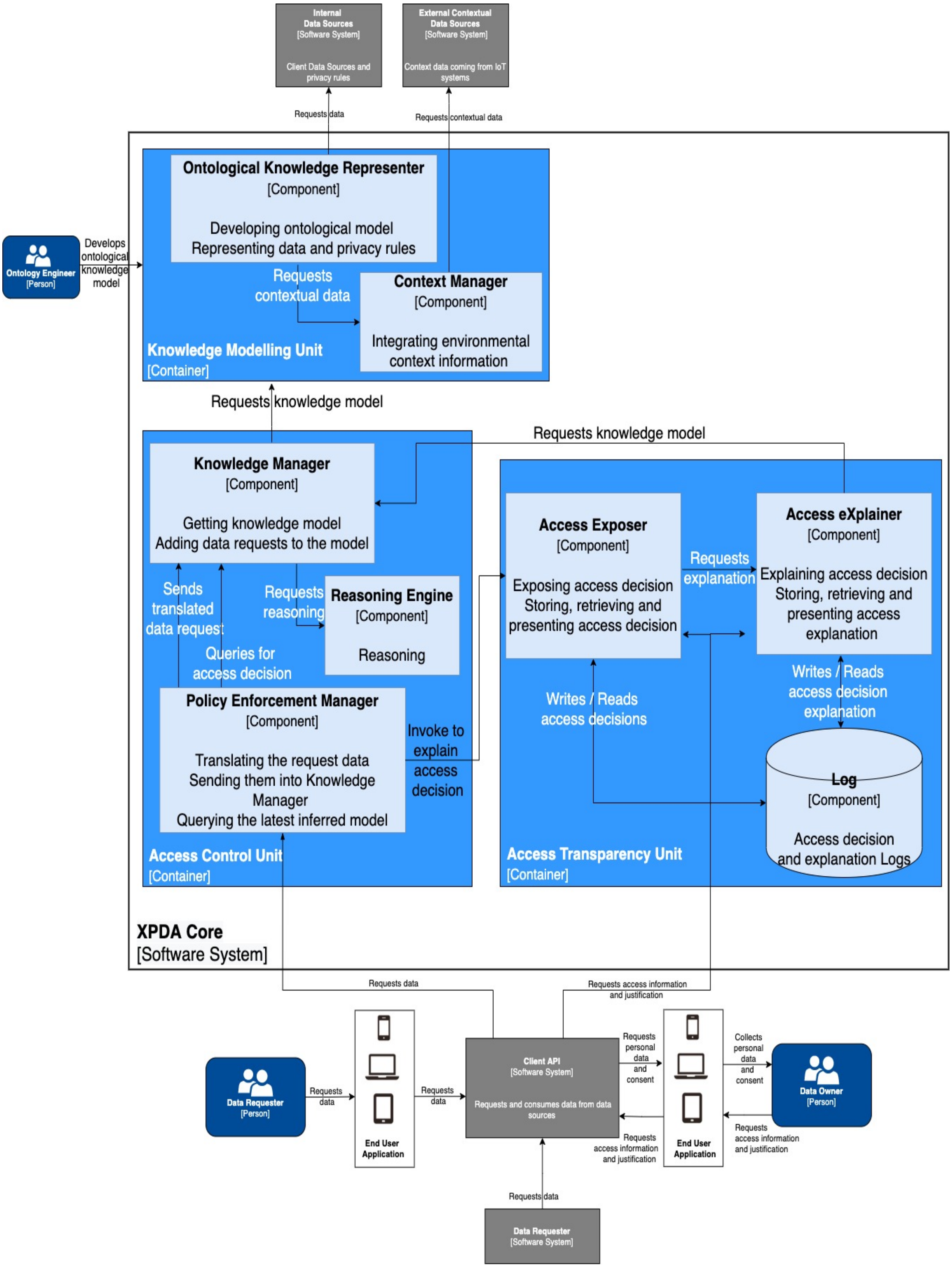

Internal
Data Sources
[Software System]
Client Data Sources and privacy rules
External Contextual
Data Sources
[Software System]
Context data coming from IoT systems
Requests data
Requests contextual data
Ontology Engineer
[Person]
Develops ontological knowledge model
Ontological Knowledge Representer
[Component]
Developing ontological model
Representing data and privacy rules
Requests contextual data
Context Manager
[Component]
Integrating environmental context information
Knowledge Modelling Unit
[Container]
Requests knowledge model
Requests knowledge model
Knowledge Manager
[Component]
Getting knowledge model
Adding data requests to the model
Sends translated data request
Requests reasoning
Reasoning Engine
[Component]
Reasoning
Queries for access decision
Policy Enforcement Manager
[Component]
Translating the request data
Sending them into Knowledge Manager
Querying the latest inferred model
Invoke to explain access decision
Access Control Unit
[Container]
Access Exposer
[Component]
Exposing access decision
Storing, retrieving and presenting access decision
Requests explanation
Access eXplainer
[Component]
Explaining access decision
Storing, retrieving and presenting access explanation
Writes / Reads access decisions
Writes / Reads access decision explanation
Log
[Component]
Access decision and explanation Logs
Access Transparency Unit
[Container]
XPDA Core
[Software System]
Requests data
Requests access information and justification
Data Requester
[Person]
Requests data
End User Application
Requests data
Client API
[Software System]
Requests and consumes data from data sources
Requests personal data and consent
Requests access information and justification
End User Application
Collects personal data and consent
Requests access information and justification
Data Owner
[Person]
Requests data
Data Requester
[Software System]

Figure 3.3. Component diagram of XPDA

3.6 Sequence Diagram of XPDA

To understand the detailed functionality of XPDA, Figure 3.4 illustrates a sequence diagram to demonstrate how containers of XPDA interact with each other to complete a process of deciding for a particular data access request from a requesting entity providing required information for enhancing the transparency.

When a requesting entity, including a user of an omnichannel application or another third-party software system/service, needs to access/process the data, a request is sent to ***ACU*** through ***Client API***. After receiving the request, ***PEM*** converts the request into the same knowledge representation language and sends this representation to ***KM***. ***KM*** requests the latest knowledge model from ***KMU*** as soon as getting the request. In ***KMU***, the latest representation of domain knowledge with its current assertions and the representation of privacy rules are collected from internal and external data resources into ***KR*** due to this request. Although there are some approaches to automatically mapping the representation of rational and non-rational databases to compatible representation format [155] [156], manual tuning of this instantiation is conducted by ***Ontology Engineer*** to ensure the correctness of representations. ***CM*** is playing as a mediator to transfer this data to ***KR*** if data needs to be gathered from external data resources. After gathering the most updated knowledge in ***KR***, it is transferred to ***ACU*** and located in ***KM***, where the access request representation is also appended to the knowledge model. This version of the knowledge model is sent to ***ATU*** as a reference model for justification. Simultaneously, ***KM*** invokes ***RE*** for knowledge induction. After inference, the knowledge model is queried by ***PEM*** for access decision corresponding to the request.

While the access decision is informed to the requester entity, the access decision's detail is forwarded to ***ATU***. This information is transformed into an understandable presentation for non-expert individuals in ***AE*** and stored in ***Log***. Meanwhile, these details are dispatched to ***AX***, where it justifies the corresponding decision according to the latest knowledge model already sent through. This justification should present in a form that can be interpretable for the data subject who seeks the reason for this data access. The explanation also is added to ***Log*** for further references.

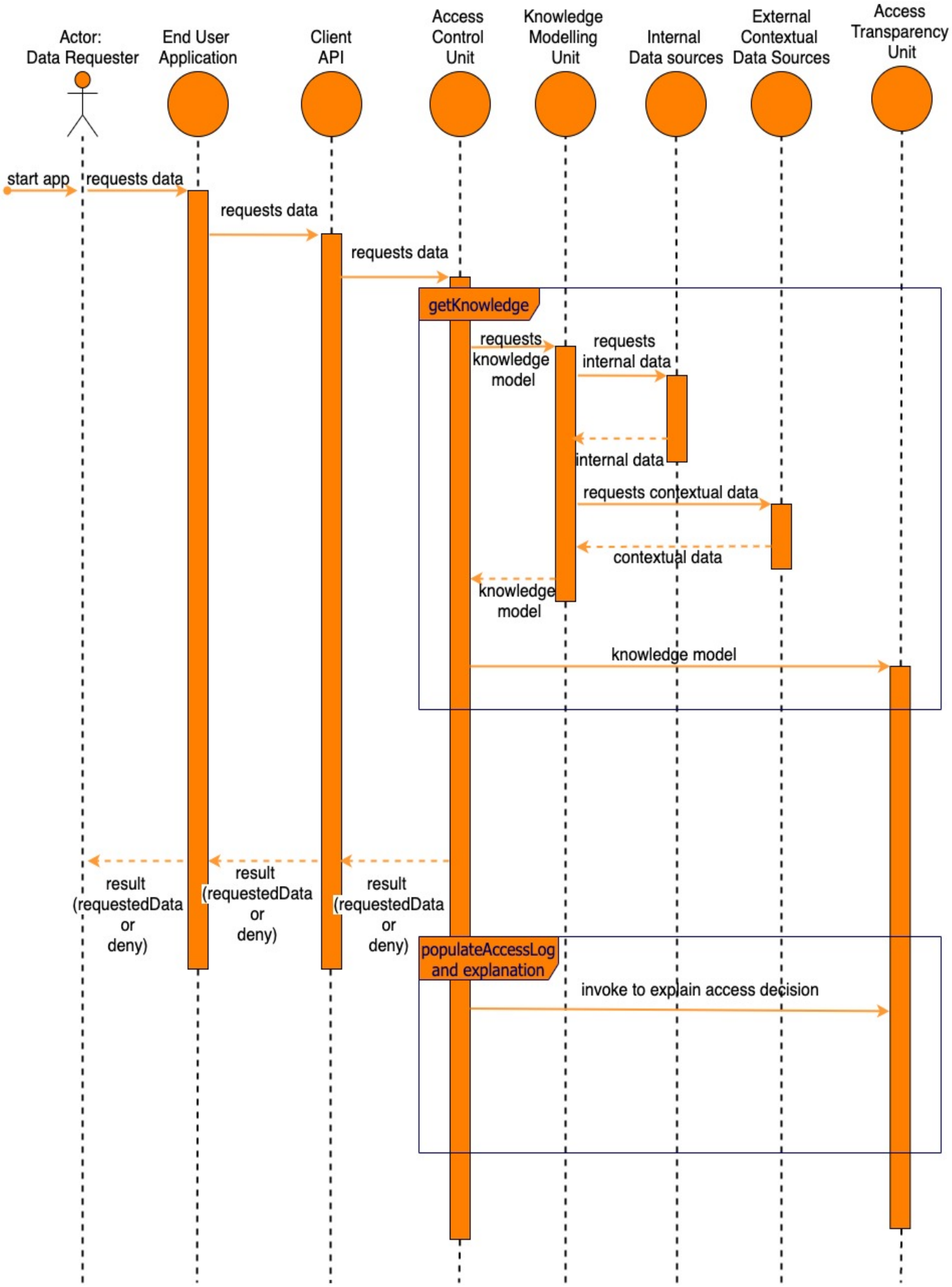


Figure 3.4. Sequence diagram of XPDA for data access request

Likewise, another sequence diagram in Figure 3.5 demonstrates the interaction of a data owner or any controlling proxy on behalf of him/her with XPDA, whenever s/he wants to attain the right of transparency on his/her data access. In this case, whenever a ***Data Owner*** is seeking the details of access information and its explanation, s/he sends a request through an ***End User Application***. This request is dispatched to ***ATU*** through the ***Client API***. Then, the corresponding information can be retrieved from ***Log*** database in ***ATU*** and presented back to ***Data Owner*** through the same channel.

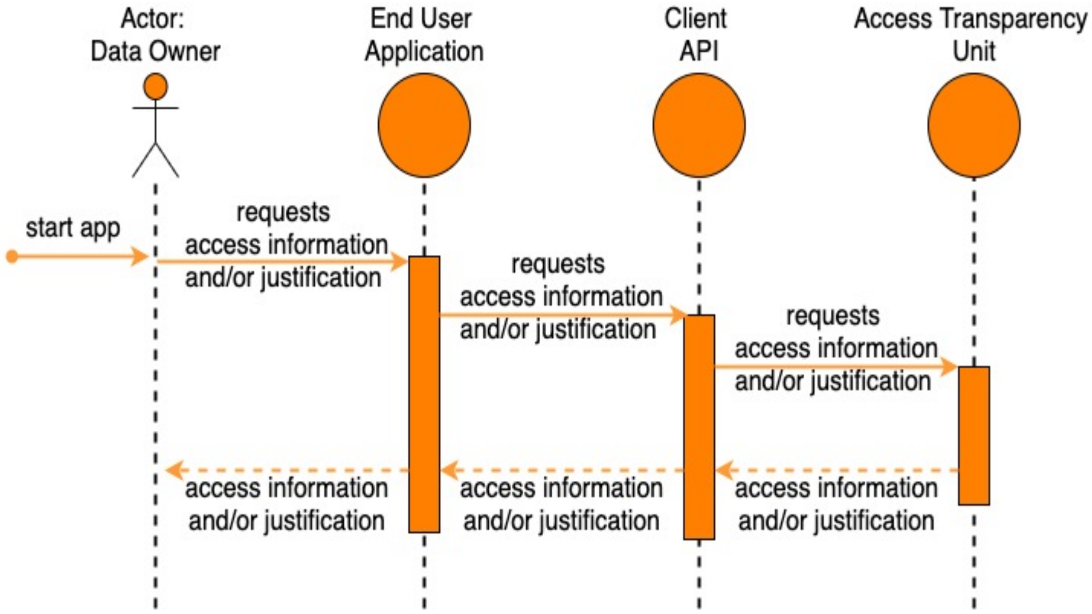


Figure 3.5. Sequence diagram of XPDA for seeking access information and explanation

3.7 Discussion

As shown in the previous sections of this chapter, the architecture of XPDA is designed to grant or deny a particular data access request based on contextual information of the system. Deploying Semantic Web as a core technology behind XPDA facilitates solving intrinsic heterogeneity of stakeholders and users of access control systems and their disparate access control criteria by providing an identical approach to convey the semantics of these criteria. It also provides sufficient flexibility to apply XPDA in different scenarios with few or no changes. Likewise, XPDA provides adequate scalability to deal with vast numbers of resources, access policies, systems, clients and attributes by putting a human intelligence (such as domain experts and ontology engineers) in the loop to precisely define a diverse range of concepts, identify their common group and specify their relation in an interoperable manner. It utilises a variety of context information not only to cover the internal context, such as user-centric attributes but also to consider the external context that can affect decision making. Exploiting ontology-based modelling to express context information through a wide range of standardised relations can provide more adaptability and interoperability with the ontologies already developed in diverse domains. In this perspective, the changes in the context can trigger the evaluation process of applicable privacy policies. When a specific context is situated, instantaneously relevant decisions can be made due to this context. Consequently, access control can act more efficiently.

According to the categorisation parameters discussed in Section 2.4.1, the expose of personal data access and presenting the reason per each access in a human-readable format in XPDA can be considered as an ex-post transparency enhancement technology which not only provides data subjects with visibility about their personal data access but also lets them know the implication of their consents on their data access. Comprehension of this implication may allow data subjects to change and improve their behavioural pattern of data disclosure. Although data subjects are considered to be the primary audience of XPDA, auditors can adapt it and use it smoothly. Further, XPDA is designed to be applied in both solitary and participatory environments. Since data subjects can only see the result of the system; therefore, it is placed in the category of passive read-only TETs. It is also designed so that its delivery mode can be implemented in both a Pull and Push Mode.

3.8 Conclusion

This chapter described the architecture of the proposed XPDA service to cater to data subjects with more control over their data by providing understandable information about the way and the reason for their personal data access. This architecture has been designed and exhibited in a hierarchy of abstraction following the C4 model.

The best practices of reviewed existing access control models have been adopted by carefully considering the deployment of context awareness and combined use of ontology constraints and rule-based approach for policy specification, enforcement and evaluation in ***Access Control Unit*** of the architecture.

Moreover, the proposed architecture has fulfilled the lack of policy awareness and privacy implication by exposing the details of the access decision and explaining the justification of this decision in its ***Access Transparency Unit***.

Consequently, the design of XPDA architecture satisfies RO2 of this research. We plan to expand this architecture by making it more interactive to data subjects, allowing them to change their privacy preferences as they comprehend the current implications. This improvement can help to privilege “the right to be forgotten” [25].

There is a study [157] published as this thesis was finalising and aimed to facilitate the “right to be informed” [25] by enhancing the ex-ante transparency. Another potential strand for future works can be integrating such approaches in the architecture of XPDA.

4 Implementation

4.1 Introduction

A prototypical implementation of the XPDA service on a motivating scenario in the health domain is discussed in this chapter. A simple arbitrary scenario of personal data access occurring in different contexts is considered as a running example to point out challenges in preserving the privacy of users in the health domain.

Besides privacy concerns of people about their health data [158], some peculiarities of the health domain make it interesting to investigate the need for more transparency about data collection, access and usage in this domain. One of the main peculiarities is that the data concerning the health of patients is often not created or edited by the patients themselves but by other subjects, such as physicians or healthcare professionals and accessed without the knowledge of the patients [159]. As a consequence, the process of disclosure of data is not as evident as in other domains. There is no precise moment when the data disclosure occurs within the medical systems, as it depends on when a patient visits the hospital or schedules an examination. However, regardless of how a patient's data reaches the system or how it is used, regulations like the GDPR are in place to protect patients' rights. GDPR considers health data as a "special category of personal data" which merit higher protection and should be processed for health-related purposes only where necessary to achieve those purposes for the benefit of natural persons and society as a whole [25]. According to Article 9 of GDPR, the processing of health data is possible if the data subject has given explicit consent to the processing of them for one or more specified purposes [25]. Also, additional strict rules should provide harmonised conditions for the processing of health data regarding specific needs. An obvious example is the processing of health data for health-related research purposes [25].

Having clear privacy policies to provide such exceptional protection measures is insufficient to remedy the patient's privacy concerns [159]. Transparency promotes the availability of alternatives for patients to verify that the system is taking or has taken the necessary precautions to protect their data. Patients must be able to check whether the agreed-upon privacy policy has been enforced. They should be able to identify the unwanted information flow [159].

To properly control access and processing of special categories of personal data such as health data, we claim the need for a more comprehensive approach that exploits not only identity and role information of the data requester but also other contextual information, such as location and time. In particular, we believe that it may improve transparency for the data subject if the access control policies for his/her health data are defined according to the current conditions of the requester and the circumambient environment, i.e. the current context. For instance, access to a patient's data should be granted to a nurse who is not only working in the hospital but also is on duty on the date which patient is hospitalised and also assigned to the patient. The consolidation of access control with contextual information is an example of an active access control model that is aware of the context of ongoing activity in providing access control.

Also, the exploitation of context as a mechanism for grouping policies and evaluating applicable ones increases policy specification reuse, eases policy update and revocation, and simplifies access control management. Therefore, our example scenario in the health domain, where contextual conditions frequently change, would show the merits of the context-based access control approach by providing more flexible, effective, and understandable preserving patient's privacy [102].

Another difficulty in a dynamic environment like health is that it is impossible to define all necessary policies for all possible situations in advance. For example, in many cases, a medical doctor can request consultancy of other specialist or request diagnostic support tests that require sharing patients' health data. A semantic-based approach can deal with an unexpected situation deploying policy adaptation, which provides the reasoning features needed to deduce new information from existing knowledge.

With regards to the considerations and prerequisites mentioned above, the following scenario is defined as a running example throughout the rest of this thesis:

> *Bob is a 25-year-old man living in Dublin. Last year, he visited the Royal Hospital for a comprehensive health screening. The Royal Hospital collected his health data (Diseases, Treatments, Allergies, ...) and demographics (Name, Gender, Date of birth, ...) with his explicit consent. This consent gave the Royal Hospital the right to access and use Bob's data for his medical care as well as in health-related collaborative researches. Under this consent, all health professionals who work in the Royal Hospital could access Bob's personal data if and only if both parties, Bob and the corresponding health professional, be located in the hospital. The Royal Hospital has had even more restricted rules for*

accessing patient's data by its staff. For example, only a nurse who is assigned to Bob can access to patient's data, or staff in various roles can access the different level of patient's data.

In this consent form, Bob also gave consent to the researchers of any research centres who are conducting collaborative research with the Royal Hospital.

Later, ADAPT (a research institute) and the Royal Hospital have collaborated in a research project named Project One to assess the allergy rate of patients at the hospital. A researcher from ADAPT, named Ramisa, has assigned the project to analyse patients' allergic data and consequently access Bob's Allergy data.

After a while, Bob revisited the Royal Hospital due to feeling pain in his chest, and he was hospitalised for half a day. Ruth and Mary, as nurses, were on duty that day at the Royal Hospital. Ruth was assigned to Bob for caring for him. Dr. Eric, a General Practitioner in Royal Hospital, visited Bob and referred him to the hospital's medical imaging unit for an X-ray. Before taking the X-ray, Tom, a radiologist at the imaging unit, checked Bob's weight from his health record. After taking the X-ray, Tom sent the X-ray report to his supervisor, Dr. Edvard, a consultant radiologist.

The remaining of this chapter describes how XPDA architecture and its corresponding building blocks are implemented. It is also illustrated how this prototype can help patients like Bob to find out more detail about their data access.

4.2 Knowledge Modelling Unit

After identifying the main concepts, their types and relations between them in the motivating scenario, an ontology is specified and represented by Web Ontology Language (OWL2) in Protégé ontology editor and knowledge base framework [160]. The graph representation for the ontology is demonstrated in Figure 4.1.

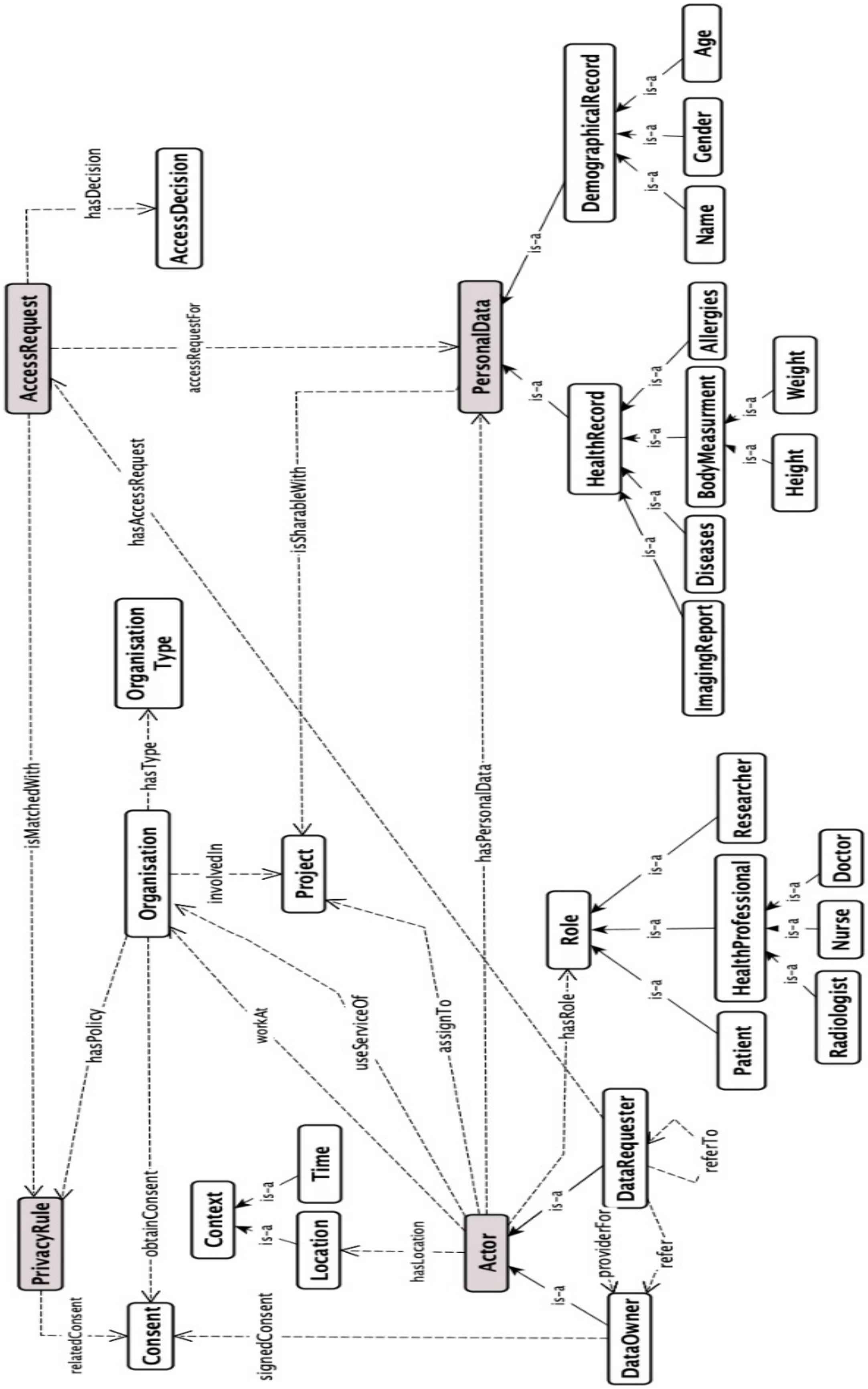


Figure 4.1. Graph representation of the ontological knowledge model for the motivating scenario

As discussed earlier in the previous chapter, XPDA service, as a context-centric approach, treats context as a first design principle in its knowledge representation. It adopts and deploys a widely accepted definition of context across several definitions collected from various areas of research [161] as “any information that can be used to characterise the situation of an entity” [162]. Therefore, the context acts as a mediator between the entities requiring access to data resources and the set of access decisions assigned to these resources in XPDA. In this prototype, two different types of contexts are considered for entities of the ontology as follows:

- Actor Context: There are two entities defined as subsumed classes of the ***Actor*** class as follows:
  - ***Data Requester*** is defined as an authenticated individual seeking permission to access or use ***Personal Data***.
  - ***Data Owner*** is defined as an individual to whom personal data is related.

  Actor context defines the specific contexts that must be held or exercised by an actor in order to control (in the case of ***Data Owner***) or obtain rights to access (in the case of ***Data Requester***) to ***Personal Data***. Actor contexts in this scenario are included as:
  - ***Role***; is assigned to an ***Actor*** based on his/her job function within the organisation (in the case of ***Data Requester***) or state in the health domain (in the case of ***Data Owner***). A role hierarchy is defined as sub-sumed classes in order to specify the role-permission assignment privacy rules.
  - ***Location;*** represents the abstraction of a physical location.
- Environment Context: Operational and situational conditions such as date and time are considered as environment context and defined as data properties of ***Access Request***. They are not associated with a particular ***Actor*** or a specific ***Personal Data*** but may nonetheless be relevant in applying a ***Privacy Rule***.

Therefore, an ***Access Request*** to certain types of ***Personal Data*** about ***Data Owner,*** who is holding a pre-defined ***Role*** as a ***Patient,*** can be provided by a given ***Data Requester,*** with a specific ***Role*** in a particular type of ***Organisation.*** This ***Access Request*** relies on the particular ***Location*** of both ***Actor***s. An ***Access Decision*** can be made by evaluating a given ***Access Request*** against a set of ***Privacy Rule***s to make an

Access Decision if a ***Consent*** signed by ***Data Owner*** agrees to disclose that particular ***Personal Data.***

Different circumstances of privacy rule specification and enforcement on various kinds of represented knowledge of the motivating scenario are considered as deployment settings. These settings are defined through three different use cases where the service aims to provide a human-readable explanation for corresponding patients' data access in each use case. Details of these use cases are illustrated in Table 4.1, while the concise deployment settings are demonstrated in Table 4.2.

This prototype uses SWRL [34] to specify privacy rules through SWRL plugin in the Protégé-OWL ontology development toolkit. In SWRL, the antecedent (called the body) and the consequent (called the head) are defined as OWL classes, properties and individuals. Therefore, the antecedent encodes the conditions specified in the privacy rule, whereas the consequent encodes the implied access decision. Table 4.3 depicts the full set of privacy rules defined for the motivating scenario and their equivalent in SWRL.

The ***Context Manager*** performs context processing to integrate environmental context information and formalised them in the format of OWL assertion. In order to keep this prototype easy to implement and understand, the functionality of ***CM*** is not included in the prototype. The contextual information that needs to collect from external resources, such as location, is provided and appended to the ontology manually.

Table 4.1. Uses cases of the motivating scenario

Use-case 1	*Bob is a 25 years old man living in Dublin. Last year, he visited the Royal Hospital for a comprehensive health screening.* *The Royal Hospital collected his health data (Diseases, Treatments, Allergies, ...) and demographics (Name, Gender, Date of birth, ...) with his explicit consent.* *After a while, Bob visited the Royal Hospital again due to feeling a pain in his chest and he was hospitalised for half a day.* *Ruth and Mary, as nurses, were on duty that day at the Royal Hospital. Ruth was assigned to Bob for caring for him.*
Use-case 2	*Bob is a 25 years old man living in Dublin. Last year, he visited the Royal Hospital for a comprehensive health screening.* *The Royal Hospital collected his health data (Diseases, Treatments, Allergies, ...) and demographics (Name, Gender, Date of birth, ...) with his explicit consent.* *After a while, Bob visited the Royal Hospital again due to feeling a pain in his chest and hospitalised for half a day.* *Dr. Eric, a General Practitioner in Royal Hospital, visited Bob and referred him to the hospital's medical imaging unit for an X-ray. Before taking the X-ray, Tom, a radiologist at the imaging unit, checked Bob's weight from his health record. After taking the X-ray, Tom sent the X-ray report to his supervisor, Dr. Edvard, a consultant radiologist.*
Use-case 3	*Bob is a 25 years old man living in Dublin. Last year, he visited the Royal Hospital for a comprehensive health screening.* *The Royal Hospital collected his health data (Diseases, Treatments, Allergies, ...) and demographics (Name, Gender, Date of birth, ...) with his explicit consent.* *Later, ADAPT (a research institute) and the Royal Hospital have collaborated in a research project named Project One to assess the allergy rate of patients at the hospital. A researcher from ADAPT, named Ramisa, has assigned the project to analyse patients' allergic data and consequently access to Bob's Allergy data.*

Table 4.2. Deployment settings for use cases of the motivating scenario

	Access Decision Circumstances	
	Involved Knowledge Type (axioms)	**Involved Rules Number/Type**
Use-case 1	Asserted	Single
Use-case 2	Asserted + Inferred	Single
Use-case 3	Asserted + Inferred	Multiple - Nested

Table 4.3. Full set of specified privacy rules of the motivating scenario and their equivalent in SWRL

Rule #1	Description	A nurse assigned to a patient of a health care centre has access to the information regarding the disease of the patient - only when both are on location.
	SWRL	DataRequester(?dr) ^ DataOwner(?do) ^ AccessRequest(?ar) ^ Diseases(?dis) ^ Patient(?pt) ^ Organisation(?org) ^ Location(?loc) ^ Nurse(?nur) ^ providerFor(?dr, ?do) ^ hasRole(?dr, ?nur) ^ hasAccessRequest(?dr, ?ar) ^ accessRequestFor(?ar, ?dis) ^ hasPersonalData(?do, ?dis) ^ hasRole(?do, ?pt) ^ workAt(?dr, ?org) ^ hasLocation(?dr, ?loc) ^ hasLocation(?do, ?loc) ^ Consent(?con) ^ signedConsent(?do, ?con) ^ obtainConsent(?hcc, ?con) -> hasDecision(?ar, grant) ^ isMatchWith(?ar, pr-1)
Rule #2	Description	Patient body measurement data can be accessed by a health professional.
	SWRL	DataOwner(?do) ^ hasRole(?do, ?pt) ^ Patient(?pt) ^ hasPersonalData(?do, ?pd) ^ BodyMeasurement(?pd) ^ HealthProfessional(?hltProf) ^ DataRequester(?dr) ^ hasRole(?dr, ?hltProf) ^ hasAccessRequest(?dr, ?req) ^ AccessRequest(?req) ^ accessRequestFor(?req, ?pd) ^ Consent(?con) ^ signedConsent(?do, ?con) -> hasDecision(?req, grant) ^ isMatchWith(?req, pr-2)
Rule #3	Description	A researcher is assigned to all projects the associated research institute is involved in.
	SWRL	DataRequester(?dr) ^ hasRole(?dr, ?rsc) ^ Researcher(?rsc) ^ Organisation(?resIns) ^ hasType(?resIns, researchInstitute) ^ workAt(?dr, ?resIns) ^ Project(?prj) ^ involvedIn(?resIns, ?prj) -> assignTo(?dr, ?prj)
Rule #4	Description	With consent - the personal health data of a patient in a health care centre may be shared with any projects which said health care centre is involved.
	SWRL	DataOwner(?do) ^ hasRole(?do, ?pt) ^ Patient(?pt) ^ PersonalData(?pd) ^ hasPersonalData(?do, ?pd) ^ Consent(?con) ^ signedConsent(?do, ?con) ^ Organisation(?hcc) ^ hasType(?hcc, healthCareCentre) ^ useServiceOf(?do, ?hcc) ^ obtainConsent(?hcc, ?con) ^ Project(?prj) ^ involvedIn(?hcc, ?prj) -> isShareableWith(?pd, ?prj)
Rule #5	Description	A researcher assigned to a project can assess the data shared within the confines of said project.
	SWRL	DataRequester(?dr) ^ hasRole(?dr, ?rsc) ^ Researcher(?rcs) ^ Project(?prj) ^ assignTo(?dr, ?prj) ^ AccessRequest(?ar) ^ hasAccessRequest(?dr, ?ar) ^ accessRequestFor(?ar, ?pd) ^ PersonalData(?pd) ^ isShareableWith(?pd, ?prj) -> hasDecision(?ar, grant) ^ isMatchWith(?ar, pr-5)

4.3 Access Control Unit

In ***ACU***, ***KM*** interacts with other components to implement access policy enforcement and evaluation. It keeps track of dependencies between ontologies, such as import relationships. It retrieves ontologies from identifying URIs, either locally or remotely, after getting the representation of domain knowledge from ***KMU*** as an OWL file. It uses the Apache Jena ontology API[8] to load ontological models with the information shaped in the ontologies and policies.

Whenever the ***PEM*** receives an access request from a data requester, it translates the request data into OWL assertion and sends it into ***KM*** for updating domain knowledge. Then ***KM*** invokes the ***RE*** to perform reasoning with the updated model. ***RE*** is the key component of the ***ACU***, responsible for getting asserted fact about the current state, checking the consistency of current state assertions, and implementing the actual inference to provide new knowledge for policy evaluation. In this prototype, ***RE*** is implemented with the Pellet reasoner [35]. Pellet instance contains a repository called Terminological Box (TBox), which stores axioms that describe concepts in the ontology, relations between them and their hierarchy, as well as a repository for axioms describing the current relationship between concepts and their instantiated individuals called the Assertional Box (ABox). It is noticeable that within an access control session, the TBox remains immutable and is stored in a local cache within the OWL-API, while the ABox, which might change at each access request, need to be reloaded at the evaluation time. Table 4.4 depicts asserted axioms and inferred axioms corresponding to the Use-case 3 of the motivating scenario. Access decision can be identified due to applying an appropriate SPARQL query to the latest inferred model. Table 4.5 illustrates the SPARQL query applied to identify all granted data access in the motivating scenario and its result. If it is induced to grant access to the request, then the request's details and its corresponding decision are sent out to ***ATU***.

[8] https://jena.apache.org/documentation/ontology/

Table 4.4. Asserted and inferred axioms corresponding to the Use-case 3 of the motivating scenario

Rule Enforcement

Assigning researcher to the research project	DataRequester(?dr) ^ hasRole(?dr, ?rsc) ^ Researcher(?rsc) ^ Organisation(?resIns) ^ hasType(?resIns, researchInstitute) ^ workAt(?dr, ?resIns) ^ Project(?prj) ^ involvedIn(?resIns, ?prj) -> assignTo(?dr, ?prj)
Sharing health data to the research purpose	DataOwner(?do) ^ hasRole(?do, ?pt) ^ Patient(?pt) ^ PersonalData(?pd) ^ hasPersonalData(?do, ?pd) ^ Consent(?con) ^ signedConsent(?do, ?con) ^ Organisation(?hcc) ^ hasType(?hcc, healthCareCentre) ^ useServiceOf(?do, ?hcc) ^ obtainConsent(?hcc, ?con) ^ Project(?prj) ^ involvedIn(?hcc, ?prj) -> isShareableWith(?pd, ?prj)
Accessing researcher to data	DataRequester(?dr) ^ hasRole(?dr, ?rsc) ^ Researcher(?rcs) ^ Project(?prj) ^ assignTo(?dr, ?prj) ^ AccessRequest(?ar) ^ hasAccessRequest(?dr, ?ar) ^ accessRequestFor(?ar, ?pd) ^ PersonalData(?pd) ^ isShareableWith(?pd, ?prj) -> hasDecision(?ar, grant) ^ isMatchWith(?ar, pr-5)

Asserted Axioms

dtR-3 : DataRequester
rsch-1: Researcher
prj-1: Project
org-2 : Organisation
dtO-1 : DataOwner
pt-1 : Patient
alg-1 : Allergies
con-1 : Consent
org-1 : Organisation
req-3 : AccessRequest
pr-5 : PrivacyRule
<dtR-3, rsch-1> : hasRole
<Org-2, researchInstitute> : hasType
<dtR-3, org-2> : workAt
<org-2, prj-1>: involvedIn
<dtO-1, pt-1> : hasRole
<dtO-1, alg-1> : hasPersonalData
<dtO-1, con-1> : signedConsent
<Org-1, healthCareCentre> : hasType
<dtO-1, org-1> : useServiceOf
<org-1, con-1> : obtainConsent
<org-1, prj-1> : involvedIn
<dtR-3, req-3> : hasAccessRequest
<req-3, alg-1> : accessRequestFor

Inferred Axioms

alg-1: PersonalData

<dtR-3, prj-1> : assignedTo

<alg-1, prj-1> : isShareableWith

<req-3, grant> : hasDecision

<req-3, pr-5> : isMatchWith

Table 4.5. SPARQL query to identify all granted data access in the motivating scenario

SPARQL Query

```
PREFIX ns: <http://www.semanticweb.org/XPDA#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?GrantedAccessRequest ?DataRequester ?AccessedDataName  ?DataOwner

WHERE {
        ?AccessRequest rdf:type ns:AccessRequest.
        ?AccessRequest rdfs:label ?GrantedAccessRequest.
        ?AccessRequest ns:hasDecision ns:grant.
        ?AccessRequest ns:accessRequestFor ?AccessedData.
        ?AccessedData rdf:type ns:PersonalData.
        ?AccessedData rdfs:label ?AccessedDataName.

        ?dtO rdf:type ns:DataOwner .
        ?dtO ns:hasPersonalData ?AccessedData.
        ?dtO ns:hasPersonalData ?demO.
        ?demO rdf:type ns:DemographicalRecord.
        ?demO ns:name ?DataOwner.

        ?dtR rdf:type ns:DataRequester .
        ?dtR ns:hasAccessRequest ?AccessRequest.
        ?dtR ns:hasPersonalData ?demR .
        ?demR rdf:type  ns:DemographicalRecord.
        ?demR ns:name ?DataRequester .
}
```

Result

```
-----------------------------------------------------------------------------------------
| GrantedAccessRequest | DataRequester | AccessedDataName | DataOwner |
=========================================================
| "Access request-1"      | "Ruth"          | "Disease"             | "Bob"       |
| "Access request-2"      | "Tom"           | "Weight"              | "Bob"       |
| "Access request-3"      | "Ramisa"        | "Allergy"             | "Bob"       |
-----------------------------------------------------------------------------------------
```

4.4 Access Transparency Unit

As discussed in the previous chapter, ***ATU*** should provide the data subject with transparent and interpretable details of access to his/her data.

AE needs to know which access requests have been asked and what decisions have been made according to these requests. Then, ***AE*** presents this information to ***Data Owner*** or any controller proxies on behalf of him/her in an understandable natural language form. This prototype implements these functionalities in a way that just granted access request is taking into account. Therefore, as the result of an access decision is "grant", the request's detail is sent out from ***PEM***. Then, to improve the readability, ***AE*** deploys a simple gap-filling approach to generate a template natural sentence to present this decision. All details of the granted access decision and its generated presentation are stored in ***Log*** for future retrieval. Meanwhile, the same information is passed to ***AX*** to provide an interpretable explanation of the reason for this granted access. Table 4.6 shows the sample of the data which is stored in ***Log*** by ***AE***, which contains all received information about granted accesses and the corresponding generated sentences per each access in the motivating scenario.

Table 4.6. Stored data in Log by AE

Data Requester	**Data Owner**	**Accessed Data**	**Generated Presentation**
Ruth	Bob	Disease	**Ruth** has access to **Bob**'s **Disease** data
Tom	Bob	Weight	**Tom** has access to **Bob**'s **Weight** data
Ramisa	Bob	Allergy	**Ramisa** has access to **Bob**'s **Allergy** data

AX implements two main functionalities, namely ***eXplanation Generator*** and ***eXplanation Interpreter*** as follows:

1. **eXplanation Generator (XG):** retrieves the corresponding outcome from ***AE*** and generates corresponding justification per each output using OWL Explanation API [163]. A justification is a kind of explanation for an entailment (inference), a minimal subset of the ontology that is sufficient for the entailment to hold [164]. OWL Explanation API needs to deploy the same reasoner used in ***RE*** to generate a set of justifications per each access decision. In this prototype, the first justification of the set is dispatched to the ***eXplanation Interpreter*** as an output of this component.
2. **eXplanation Interpreter (XI):** as evidenced through the results of a user study described in Section 5.4.1, explanations generated through OWL API can be very difficult or impossible to understand for a range of people, from novices to those with several years of experience with OWL. Therefore, this functionality is implemented to interpret the explanation generated through OWL Explanation API into more understandable formats for data owners. This interpretation should be deterministic, semantically equivalent to the original and easy to understand. The details of this interpretation which would generate a new explanation, can be described as follows:
    1. A corresponding template sentence generated by ***AE*** is expressed as an access decision which will be explained.

2. The following steps are performed to translate rules were included in an explanation generated through OWL Explanation API into natural languages:
   - All OWL explanation lines expressing the involved rule are identified through a pattern matching approach using regular expressions.
   - The corresponding description of the identified rule/s is retrieved through a look-up search in a table created to map a rule presented in Manchester OWL syntax [165] into its description in natural language.
3. In most of the explanations generated through OWL Explanation API, there can be axioms that are surplus or most probably would not confer the intended justification. These parts can cause various usability and understandability problems and need to be pruned [166]. In this prototype, axioms that specify general OWL constructs such as "*Range*", "*Domain*", and explicit class typing, e.g. "*Type*", are removed from the explanation. Likewise, since the details of the access request are previously depicted in the template sentence (described in Paragraph 1), the axioms which define "*hasAccessRequest*" and "*accessRequestFor*" object properties are removed from the explanation in order to reduce redundancy.
4. The remaining axioms are verbalised and converted into controlled English texts. In general, it is preferred to retain the structure of the input axiom generated through OWL Explanation API intact and keep the structure of the verbalised sentence as similar as possible to the structure of the input axiom. Therefore, each axiom is split into elements, and each element is replaced with a corresponding proper phrase which annotated to any element of the ontology using rdfs:label. The annotation is applied based on the following guidelines adopted from previous studies [167] [168]:
   - Singular proper names (preferably capitalised) and singular countable nouns denote individuals and classes, respectively.

- Object properties of the ontology are annotated with their equivalent phrases. Table 4.7 depicts some sample labels used to annotating object properties in the prototype implementation.

Table 4.7. Sample labels for annotating properties in the prototype implementation

Object property	**rdfs:Label**
useServiceOf	uses service of
hasRole	is a
hasPersonalData	has personal data
hasLocation	is at
workAt	works at
involvedIn	is involved in
signedConsent	signed
obtainConsent	obtains

5. Finally, the verbalised explanation is visualised as a diagram of the abstract graph using Graphviz API[9]. It is done via setting attributes of nodes, edges, or subgraphs in the DOT language text file [169]. A DOT file is created per each verbalised explanation and dispatched to the API, where a drawing of a graph in a graphics format is provided as an output. To create a DOT file, each verbalised axiom in the explanation is converted to a directed subgraph where its source and target node and their connected edge are defined and labelled by the subject, the object and the predicate of the axiom, respectively. Following attributes are applied to present a particular verbalised explanation as a directed graph in DOT language:
    - All nodes of subgraphs for presenting explanation axioms are drawn, by default, with an "ellipse" shape, and their edges are drawn with a solid line and normal arrowhead in black colour.
    - To make the access decision more visible, the nodes for presenting data owner and data requester are drawn with a "doublecircle" shape. In contrast, the node for presenting a data requester is styled as "rounded, filled" in a differentiated colour. Likewise, the node

[9] https://graphviz.org/documentation/

for presenting accessed data is drawn with a " Cylinder" shape and styled as "rounded, filled" in a differentiated colour. The edge for presenting an access decision is drawn as a dashed line in a differentiated colour.

All details of content per each presentation format (textual and visual), generated during the interpretation explained above, are stored in the appropriate file format in ***Log*** for future retrieval.

The set of axioms involved in composing a justification of OWL Explanation API, its equivalent verbalised explanation, along with the visualised presentation from Graphviz API for a sample data access decision per each use case of the motivating scenario are illustrated in Tables 4.8 - 4.10.

Also, Tables 4.11 – 4.13 illustrate the final explanation generated through the XPDA service in the textual and visual formats for a sample data access decision per each use case of the motivating scenario.

Table 4.8. Axioms from OWL explanation API, their equivalent verbalised explanation and visualised presentation from Graphviz API for a sample access decision of Use-case 1

OWL Explanation axioms	pt-1 Type Patient org-1 obtainConsent con-1 dtR-1 hasLocation loc-1 loc-1 Type Location nur-1 Type Nurse dtO-1 Type DataOwner dtO-1 hasPersonalData dis-1 dtO-1 hasRole pt-1 dtR-1 Type DataRequester dtR-1 hasAccessRequest req-1 obtainConsent Domain Organisation req-1 accessRequestFor dis-1 accessRequestFor Domain AccessRequest dtR-1 hasRole nur-1 dtR-1 workAt org-1 dtO-1 signedConsent con-1 dtO-1 hasLocation loc-1 dtR-1 providerFor dtO-1 dis-1 Type Diseases con-1 Type Consent
Verbalised Explanation	Royal Hospital obtains Consent Form #1 Ruth is at Royal Hospital Bob has personal data Disease Bob is a Patient Ruth is a Nurse Ruth works at Royal Hospital Bob signed Consent Form #1 Bob is at Royal Hospital Ruth assigned to Bob

DOT file	digraph G {
	“Royal Hospital” -> “Consent Form #1” [label=“obtains “];
	“Ruth” -> “Royal Hospital” [label=“is at “];
	“Bob” -> “Disease” [label=“has personal data “];
	“Bob” -> “Patient” [label=“is a “];
	“Ruth” -> “Nurse” [label=“is a “];
	“Ruth” -> “Royal Hospital” [label=“works at “];
	“Bob” -> “Consent Form #1” [label=“signed “];
	“Bob” -> “Royal Hospital” [label=“is at “];
	“Ruth” -> “Bob” [label=“assigned to “];
	“Ruth” -> “Disease” [label=“ has access to “ , fontsize=14 , fontname=“times-bold”, fontcolor=brown3 , penwidth=2 , style=dashed, color = brown3];
	“Bob” [shape= doublecircle];
	“Ruth” [fillcolor = brown3, style=“rounded,filled” , shape= doublecircle];
	“Disease” [fillcolor = brown3, style=“rounded,filled” , shape=cylinder];}

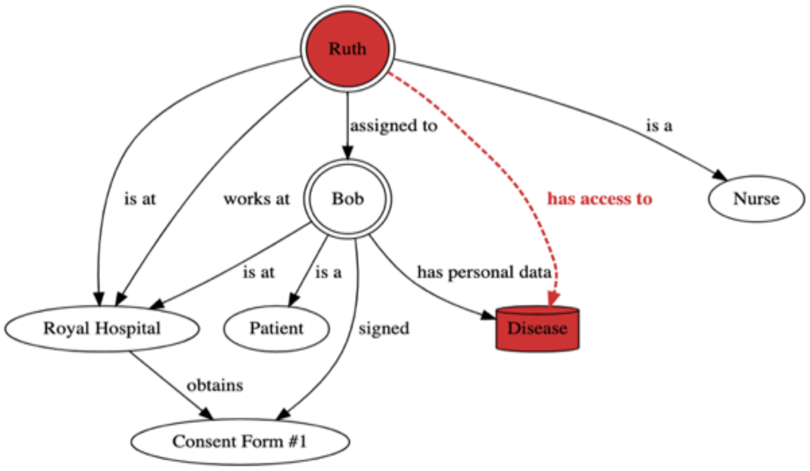
Graph	 Ruth assigned to is a is at works at Bob has access to Nurse is at is a has personal data Royal Hospital Patient signed Disease obtains Consent Form #1

Table 4.9. Axioms from OWL explanation API, their equivalent verbalised explanation and visualised presentation from Graphviz API for a sample access decision of Use-case 2

OWL Explanation axioms	pt-1 Type Patient dtR-2 hasRole rdg-1 dtO-1 Type DataOwner Weight SubClassOf BodyMeasurement dtO-1 hasRole pt-1 wgt-1 Type Weight dtR-2 Type DataRequester dtR-2 hasAccessRequest req-2 req-2 Type AccessRequest req-2 accessRequestFor wgt-1 dtO-1 hasPersonalData wgt-1 rdg-1 Type Radiologist dtO-1 signedConsent con-1 Radiologist SubClassOf HealthProfessional con-1 Type Consent
Verbalised Explanation axioms	Tom is a Radiologist Weight is a kind of BodyMeasurement Bob is a Patient Bob has personal data Weight Bob signed Consent Form #1 Radiologist is a kind of HealthProfessional

DOT file	digraph G {
	“Tom” -> “Radiologist” [label=“is a “];
	“Weight” -> “BodyMeasurement” [label=“is a kind of “];
	“Bob” -> “Patient” [label=“is a “];
	“Bob” -> “Weight” [label=“has personal data “];
	“Bob” -> “Consent Form #1” [label=“signed “];
	“Radiologist” -> “HealthProfessional” [label=“is a kind of “];
	“Tom” -> “Weight” [label=“ has access to “ , fontsize=14 , fontname=“times-bold”, fontcolor=brown3 , penwidth=2 , style=dashed, color = brown3];
	“Bob” [shape= doublecircle];
	“Tom” [fillcolor = brown3, style=“rounded,filled” , shape= doublecircle];
	“Weight” [fillcolor = brown3, style=“rounded,filled” , shape=cylinder];
	}

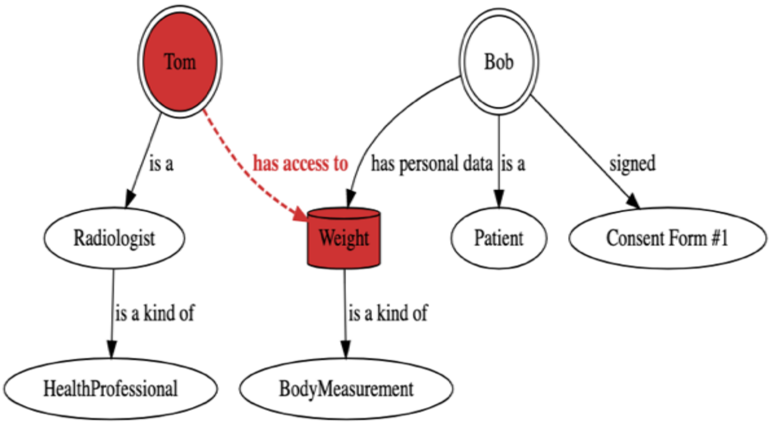
Graph	 Tom Bob is a has access to has personal data is a signed Radiologist Weight Patient Consent Form #1 is a kind of is a kind of HealthProfessional BodyMeasurement

Table 4.10. Axioms from OWL explanation API, their equivalent verbalised explanation and visualised presentation from Graphviz API for a sample access decision of Use-case 3

OWL Explanation axioms	dtR-6 workAt org-3 dtR-6 hasRole researcher dtO-1 useServiceOf org-1 org-1 hasType healthCareCentre org-3 involvedIn prj-1 dtR-6 hasAccessRequest req-7 org-1 obtainConsent con-1 org-3 hasType researchInstitute req-7 Type AccessRequest hasAccessRequest Domain DataRequester prj-1 Type Project signedConsent Domain DataOwner dtO-1 hasRole patient dtO-1 signedConsent con-1 signedConsent Range Consent req-7 accessRequestFor alg-1 dtO-1 hasPersonalData alg-1 org-1 involvedIn prj-1 hasType Domain Organisation hasPersonalData Range PersonalData
Verbalised Explanation	ADAPT is involved in Project One Bob uses service of Royal Hospital Royal Hospital is a Health Care Centre ADAPT is a Research Institute Royal Hospital obtains Consent Form #1 Ramisa works at ADAPT Bob is a Patient Royal Hospital is involved in Project One Ramisa is a Researcher Bob has personal data Allergy Bob signed Consent Form #1

DOT file	digraph G {
	“ADAPT” -> “Project One” [label=“is involved in “];
	“Bob” -> “Royal Hospital” [label=“uses service of “];
	“Royal Hospital” -> “Health Care Centre” [label=“is a “];
	“ADAPT” -> “Research Institute” [label=“is a “];
	“Royal Hospital” -> “Consent Form #1” [label=“obtains “];
	“Ramisa” -> “ADAPT” [label=“works at “];
	“Bob” -> “Patient” [label=“is a “];
	“Royal Hospital” -> “Project One” [label=“is involved in “];
	“Ramisa” -> “Researcher” [label=“is a “];
	“Bob” -> “Allergy” [label=“has personal data “];
	“Bob” -> “Consent Form #1” [label=“signed “];
	“Ramisa” -> “Allergy” [label=“ has access to “ , fontsize=14 , fontname=“times-bold”, fontcolor=brown3 , penwidth=2 , style=dashed, color = brown3];
	“Bob” [shape= doublecircle];
	“Ramisa” [fillcolor = brown3, style=“rounded,filled” , shape= doublecircle];
	“Allergy” [fillcolor = brown3, style=“rounded,filled” , shape=cylinder];}

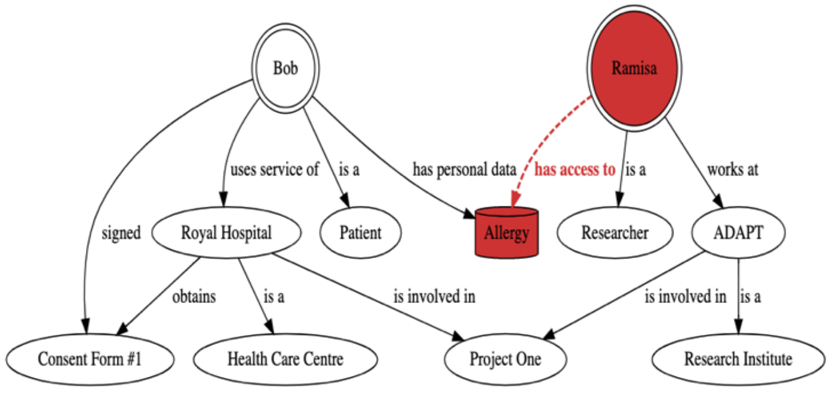
Graph	 Bob Ramisa uses service of is a has personal data has access to is a works at signed Royal Hospital Patient Allergy Researcher ADAPT obtains is a is involved in is involved in is a Consent Form #1 Health Care Centre Project One Research Institute

Table 4.11. Explanation generated by XPDA for data access in Use-case 1

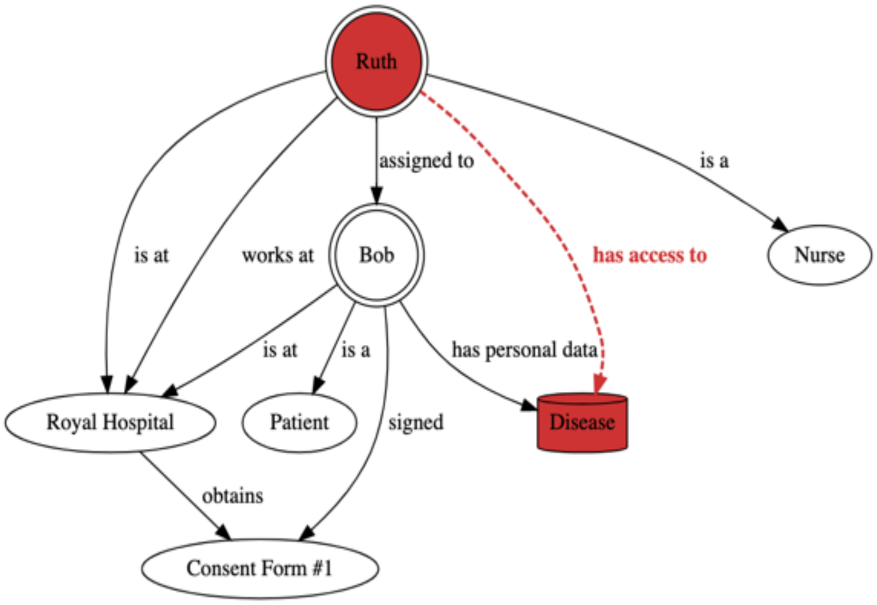
Textual format	**Personal Data Access Decision:** Ruth has access to Bob's Disease data **Explanation – the reason for access decision:** The reason of why "Ruth has access to Bob's Disease data": Based on privacy rule/s: • A nurse assigned to a patient of a health care centre has access to the information regarding the disease of the patient - only when both are on location. Matched information: Royal Hospital obtains Consent Form #1 Ruth is at Royal Hospital Bob has personal data Disease Bob is a Patient Ruth is a Nurse Ruth works at Royal Hospital Bob signed Consent Form #1 Bob is at Royal Hospital Ruth assigned to Bob
Visual format	**Personal Data Access Decision:** Ruth has access to Bob's Disease data **Explanation – the reason for access decision:** The reason of why "Ruth has access to Bob's Disease data": Based on privacy rule/s: • A nurse assigned to a patient of a health care centre has access to the information regarding the disease of the patient - only when both are on location. Matched information:  Ruth assigned to is a Nurse is at works at Bob has access to is at is a has personal data Royal Hospital Patient signed Disease obtains Consent Form #1

Table 4.12. Explanation generated by XPDA for data access in Use-case 2

<table>
<tr>
<td>Textual format</td>
<td>

Personal Data Access Decision:
Tom has access to Bob's Weight data

Explanation – the reason for access decision:
The reason why "Tom has access to Bob's Weight data":

Based on privacy rule/s:
- Patient body measurement data can be accessed by a health professional.

Matched information:
Tom is a Radiologist
Weight is a kind of BodyMeasurement
Bob is a Patient
Bob has personal data Weight
Bob signed Consent Form #1
Radiologist is a kind of HealthProfessional

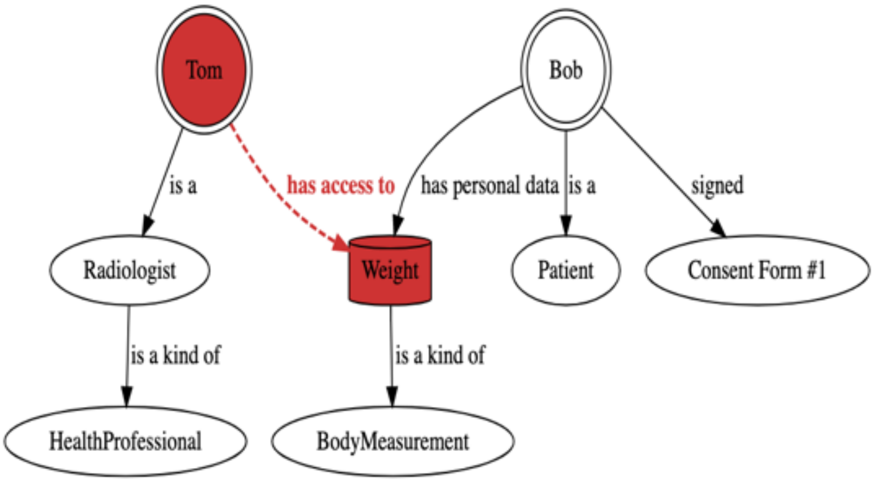
</td>
</tr>
<tr>
<td>Visual format</td>
<td>

Personal Data Access Decision:
Tom has access to Bob's Weight data

Explanation – the reason for access decision:
The reason why "Tom has access to Bob's Weight data":

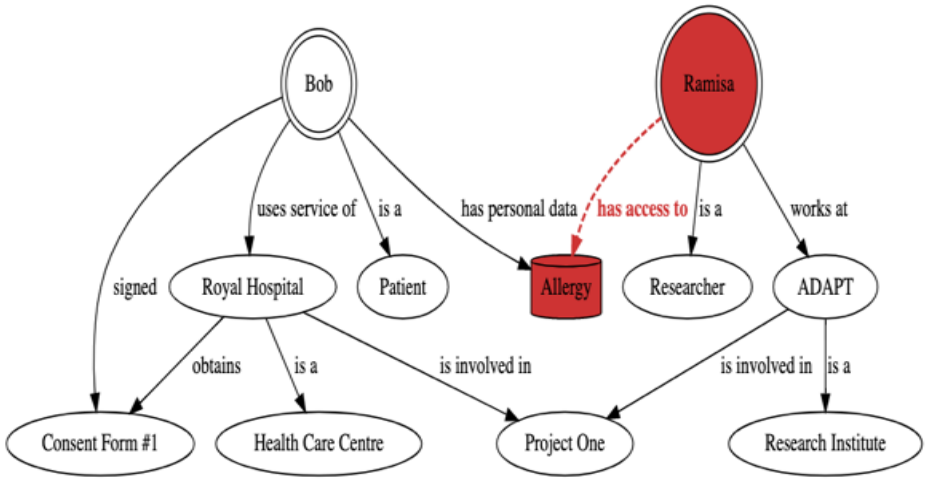
Based on privacy rule/s:
- Patient body measurement data can be accessed by a health professional.

Matched information:



</td>
</tr>
</table>

Table 4.13. Explanation generated by XPDA for data access in Use-case 3

Textual format	**Personal Data Access Decision:** Ramisa has access to Bob's Allergy data **Explanation – the reason for access decision:** The reason why "Ramisa has access to Bob's Allergy data": Based on privacy rule/s: • A researcher assigned to a project can assess the data shared within the confines of said project. • With consent - the personal health data of a patient in a health care centre may be shared with any projects which said health care centre is involved. • A researcher is assigned to all projects the associated research institute is involved in. Matched information: ADAPT is involved in Project One Bob uses service of Royal Hospital Royal Hospital is a Health Care Centre ADAPT is a Research Institute Royal Hospital obtains Consent Form #1 Ramisa works at ADAPT Bob is a Patient Royal Hospital is involved in Project One Ramisa is a Researcher Bob has personal data Allergy Bob signed Consent Form #1
Visual format	**Personal Data Access Decision:** Ramisa has access to Bob's Allergy data **Explanation – the reason for access decision:** The reason why "Ramisa has access to Bob's Allergy data": Based on privacy rule/s: • A researcher assigned to a project can assess the data shared within the confines of said project. • With consent - the personal health data of a patient in a health care centre may be shared with any projects which said health care centre is involved. • A researcher is assigned to all projects the associated research institute is involved in. Matched information:  Bob Ramisa uses service of is a has personal data has access to is a works at signed Royal Hospital Patient Allergy Researcher ADAPT obtains is a is involved in is involved in is a Consent Form #1 Health Care Centre Project One Research Institute

4.5 Discussion

This chapter described the details of a prototype implementation of XPDA architecture to illustrate how Semantic Web technologies can provide data subjects with control over their personal data by putting out the transparency on their personal data access in an interpretable manner.

Ontological knowledge modelling facilitates the determination of the concepts and their definitions and relationships comprising data access control vocabulary in the application domain. Using the user-friendly interface of freely available ontology development tools like Protégé along with putting human in the loop not only can provide more usability and efficiency in the design phase but also make it more readable and understandable for privacy and legal regulatory experts to revise the model in the early stage of its development. The expressiveness of OWL to represent the complex relations between concepts allows to specify the complex hierarchy of personal data and also to define the complicated relations in some contextual knowledge of data requester.

Integration of SWRL rule language and OWL constraints supports the definition of privacy rules and specifying the policy. Also, latent knowledge about various concepts and their relations can be derived via deploying Pellet, which supports OWL/Rule hybrid reasoning. Then, detailed information about data access can be retrieved through querying the knowledge model using SPARQL and justification for any entailment about access decision made in the service can be explained through OWL explanation. Applying a verbalisation through labelling of ontology elements can translate a complicated machine-readable explanation either to human-understandable controlled language or more visible graph presentation.

Figure 4.2 briefly depicts how Semantic Web technologies and their corresponding tools/assets were deployed in the prototype to implement each XPDA functions.

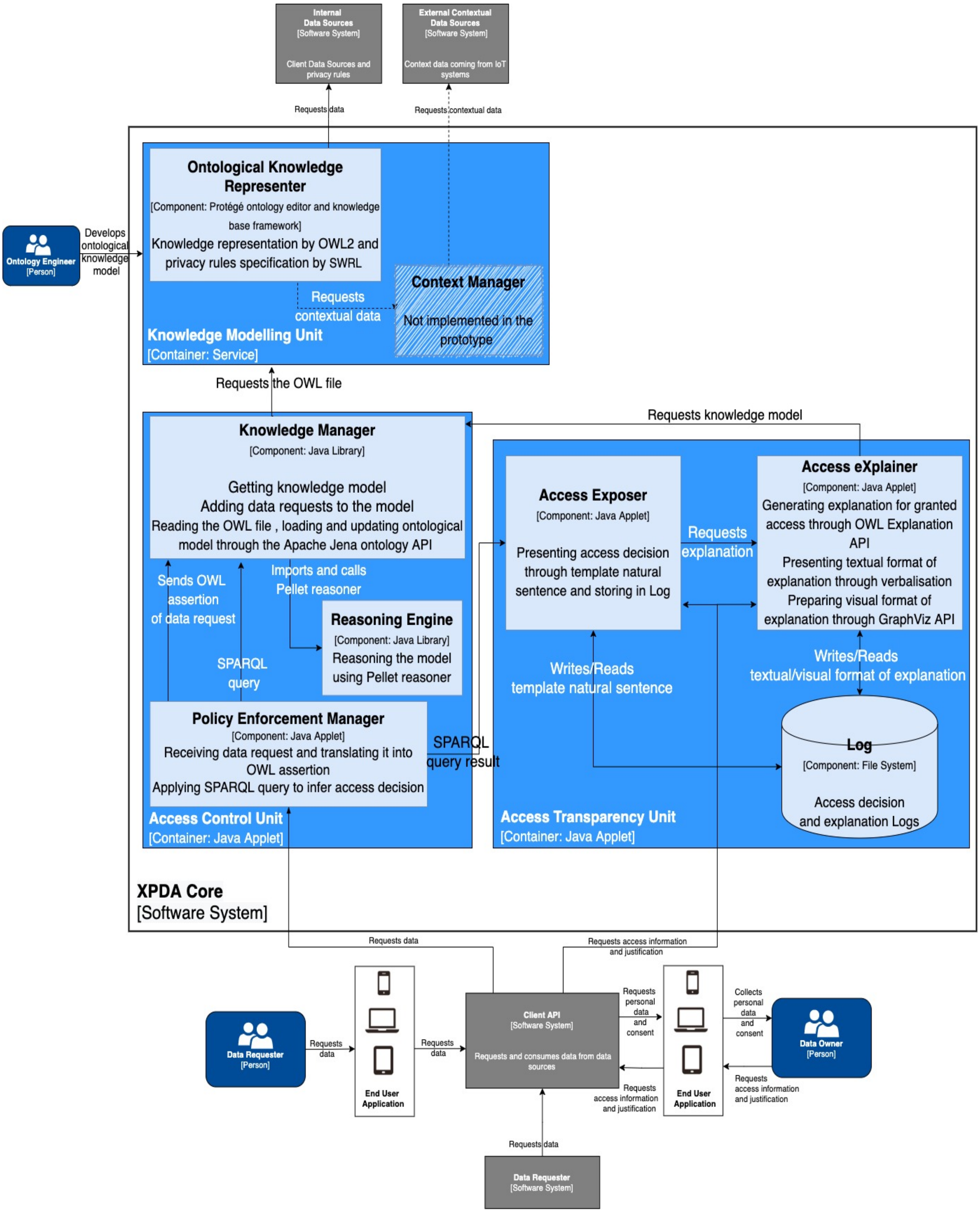

Internal
Data Sources
[Software System]
Client Data Sources and privacy rules
External Contextual
Data Sources
[Software System]
Context data coming from IoT systems
Requests data
Requests contextual data
Ontology Engineer
[Person]
Develops ontological knowledge model
Ontological Knowledge Representer
[Component: Protégé ontology editor and knowledge base framework]
Knowledge representation by OWL2 and privacy rules specification by SWRL
Requests contextual data
Context Manager
Not implemented in the prototype
Knowledge Modelling Unit
[Container: Service]
Requests the OWL file
Requests knowledge model
Knowledge Manager
[Component: Java Library]
Getting knowledge model
Adding data requests to the model
Reading the OWL file , loading and updating ontological model through the Apache Jena ontology API
Sends OWL assertion of data request
Imports and calls Pellet reasoner
Reasoning Engine
[Component: Java Library]
Reasoning the model using Pellet reasoner
SPARQL query
Policy Enforcement Manager
[Component: Java Applet]
Receiving data request and translating it into OWL assertion
Applying SPARQL query to infer access decision
SPARQL query result
Access Control Unit
[Container: Java Applet]
Access Exposer
[Component: Java Applet]
Presenting access decision through template natural sentence and storing in Log
Requests explanation
Access eXplainer
[Component: Java Applet]
Generating explanation for granted access through OWL Explanation API
Presenting textual format of explanation through verbalisation
Preparing visual format of explanation through GraphViz API
Writes/Reads template natural sentence
Writes/Reads textual/visual format of explanation
Log
[Component: File System]
Access decision and explanation Logs
Access Transparency Unit
[Container: Java Applet]
XPDA Core
[Software System]
Requests data
Requests access information and justification
Data Requester
[Person]
Requests data
End User Application
Requests data
Client API
[Software System]
Requests and consumes data from data sources
Requests personal data and consent
Requests access information and justification
End User Application
Collects personal data and consent
Requests access information and justification
Data Owner
[Person]
Requests data
Data Requester
[Software System]

Figure 4.2. Deployed Semantic Web methods and assets in the implementation of XPDA prototype

To expose personal data access and to present the reason per each access in a human-readable format in this prototype, the ex-post transparency enhancement approach of XPDA is implemented in a solitary environment in a way that data subjects can see the result of the system; therefore, it is placed in the category of Passive Read-Only. Also, since it is implemented based on query by the data subject, its delivery mode can be considered as a Pull Mode. The motivating scenario showed that XPDA could provide transparency for the reason of access to all type of data, including observed or inferred and no matter if they were either disclosed by data subjects themselves or by others. Finally, the explanation provided in this implementation can be trusted by data subjects because the correctness and completeness of the explanation provided by XPDA can be guaranteed due to the detail argued in [170]. Table 4.14 briefly illustrates the properties and functionality of the prototypical implementation of XPDA in this thesis regarding the categorisation parameters of TET mentioned in Section 2.4.

Table 4.14. Properties and functionality of XPDA regards to the categories and parameters of TET

TET category and parameter		**XPDA**	**Potential and future work**
Application Time	Ex-ante	×	✓
	Ex-post	✓	--
Target Audience	Auditors (Data Controller)	✓	--
	Data Subjects	✓	--
Environmental Context	Solitary environment	✓	--
	Participatory communities	×	✓
Interactivity Level	Passive Read-Only	✓	--
	Interactive Read-Only	×	✓
	Interactive	×	✓
Delivery Mode	Pull Mode	✓	--
	Push Mode	×	✓
Data Types Presented	Volunteered Data	✓	--
	Observed Data	✓	--
	Incidental Data	✓	--
	Derived Data	✓	--
Assurance Level	Untrusted	×	--
	Semi-Trusted	×	--
	Trusted	✓	--

4.6 Conclusion

To fulfil RO3 of this study, the undertaken approaches to implementing an XPDA service prototype on a motivating scenario in the health domain is discussed in this chapter.

The XPDA service uses Semantic Web technologies to control access over personal data of individuals based on defined privacy rules. These rules are enforced to evaluate the contextual state of an access request for making an access decision. The made decision is exposed to data subjects, and the explanation of the reason for a decision is provided in a human-readable format.

However, there are some limitations in implementing this service that need further investigation in the future. These limitations either are put in place intentionally to simplify the implementation or realised during the studies. These limitations or gaps and potential approaches for their improvement can be discussed as follows:

- One of the foremost gaps in preserving the privacy of people through ontological modelling in any related application is identified as the lack of agreed and standard vocabularies. These vocabularies not only need to describe the key characteristics of personal data and its categorisation but also should explain the purposes of handling and categories of processing to comply with the required legal bases such as consent. Although some collaborative works, discussed in Appendix A, are conducted by the author of this thesis to fulfil the gap in the time of writing this thesis, such as [43], more investigation is required to refine these vocabularies based on additional use cases and demonstrate their effectiveness in various business settings.
- In this prototype, extracting the privacy rules and specifying them in SWRL is carried out manually, which would take lots of time and labour in at-scale applications. Implementation of a component, which can automatically perform this functionality, can be considered as a potential improvement for this service.
- Likewise, the implementation of the Context Manager in the prototype was limited to manually definition of the contextual or environmental information such as location. In the future, we are planning to integrate some emerging technologies such as IoT to extract and discover contextual data in more ubiquitous environments. Also, Quality of Context (QoC) parameters [171]

can be taken into account to evaluate and ensure proper functioning in the process of adaptation in access control decision.

- The correctness and completeness of the explanation were evaluated and discussed in [170]. A user study discussed in Chapter 5 was designed and experimented with evaluating the interpretability of the explanation generated through the XPDA service. Nevertheless, a quantitative or qualitative approach needs to be proposed to evaluate its efficiency regarding system availability, including fault tolerance, load balancing, and resource consumption, mainly in terms of bandwidth and computational resources such as CPU and memory. Furthermore, since an arbitrary explanation selected across a set of justifications created per each access decision through OWL Explanation API, further investigation on the evaluation of other justifications and defining a selection method of the most appropriate one might improve the effectiveness of the service.
- The verbalisation quality depends on the linguistic features of the names used for individuals, classes, and properties in the input ontology. Therefore, probably the most visible deficiency of the described verbalisation is caused by the naming conventions used in OWL ontologies. Real-world OWL ontologies can contain complicated class, and property names do not lend themselves well to any verbalisation scheme. So, defining a more standard and precise verbalisation approach covering a broad range of names for classes and properties would be considered as another future work.
- As discussed in the previous section, the delivery mode of transparency in this prototype is implemented on a demand-basis (Pull Mode). This feature also can be implemented in a Push Mode to notify data subjects whenever any event related to their personal data access happens. This improvement can cause an increase in the interactivity level of the XPDA service, too. Table 4.14 illustrates the potential future works on the functionality of ATU concisely.

5 Evaluation

5.1 Introduction

In all the research studies, the evaluation should match the claimed contribution. Evaluation of applied work should demonstrate success in the application, while core methods should demonstrate generalisability via careful evaluation on a variety of synthetic and standard benchmarks.

As described in Chapter 1, the primary purpose of this study is to design and implement a service that exposes personal data access and explains the reason for the access in a manner that is understandable to the human user. This characteristic of generated explanation tightly aligns with either a system-centric definition of *interpretability* which is described as the "ability to explain or to present in understandable terms to a human." or a human-centred perspective of its definition as "the degree to which an observer can understand the cause of a decision" [32].

According to the literature review, most user studies in reviewed literature were refused standard approaches/measures and conducted based on arbitrary questionnaires for participants' self-expression. Also, they dismiss participants' mental model and cognition in their user study design. Both of these gaps in these user studies led to mitigating the validity of their results and their evaluation reliability.

In recent years, several models, methods, and interfaces were developed in the emerging domain of Explainable Artificial Intelligence (XAI) and Interpretable Machine Learning (IML) [172] [173] [151]. Considering the definition of XAI proposed by [174] as a "series of machine learning techniques that enables human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners", the approaches that have been taken to assess the understandability of XAI could be potentially highly relevant sources of inspiration for evaluating interpretability of our service. However, most of these works focus on computational problems and strive for functional evaluation of their approaches, while limited research effort is reported concerning their user evaluation. The need for more rigid empirical evaluation is identified in previous surveys [151] [173] [175].

Also, the lack of formal definition and standard measure for a correct or best explanation is another challenge for its empirical evaluation [176]. Even if a formal foundation exists, it does not necessarily end in practical advantage for humans as it is highly dependent on the extent to which humans can perceive it. Without proper human behaviour evaluations, it is difficult to assess this advantage for humans in practical use cases [177].

In the rest of this chapter, the background of the explanation evaluation in human behaviour, investigated in social science, will be discussed in Section 5.2. Then a preliminary taxonomy of evaluation approaches through a literature review will be presented in Section 5.3, details of experiments conducted during this study, their procedures and results will be discussed in Section 5.4.

5.2 Background of explanation evaluation in human behaviour

Experiences and findings of human behavioural studies can be used as a good source of inspiration to evaluate the extent to which an explanation of complex decision-making systems can afford interpretability to their users. The evaluation of explanation by humans has been investigated by psychologists and social scientists for decades. Within these disciplines, explanation evaluation refers to the process applied by an explainee (who receives explanation) to determine whether the explanation is satisfactory [178]. Although criteria of this satisfaction are often arbitrary and heavily influenced by cognitive biases and heuristics, as argued in the Theory of Explanatory Coherence [179], most of them are related to the way that explanation assists the explainee in understanding the underlying cause. A formal model of explanation selection based on epistemic relevance [180], as an extension to structural causal models [181], defined an explanation as a fact that, if found to be true, would constitute an actual cause of a specific event.

Foundational series of research [178] [182] [183] [184] argued that while truth, consistency and the most likely cause of a specific event are considered as essential criteria of a good explanation, it is not sufficient on its own. They demonstrated that a good explanation must be relevant to both the question and the mental model of the explainee as well. Therefore, the problem is to "resolve a puzzle in the explainee's mind about why the event happened by closing a gap in his or her knowledge" [182]. On the other hand, they proposed a conversational model of explanation and argued that explanation is not only causal attribution but also a conversation [182]. This model consists of two stages: the determination of causality in which the explainer discovers why an action/event occurred; and the social process of communicating the explanation to the explainee, which can be considered as a conversation [180].

Therefore, a good explanation needs to follow basic rules of conversation, including Grice's maxims of conversation [185]: (a) quality; (b) quantity; (c) relation; and (d) manner and should not explain any causes the explainee already knows.

Another study argued that people tend to judge the quality of explanation based on their pragmatic influence of causal behaviours [186]. This argumentation was followed by other studies to show that people assess an explanation based on its usefulness, including quality, quantity and relation [187]. In other words, while likely causes are part of good explanations, they do not strongly correlate with explanations that people find useful. Simplicity, generality, and coherence are three criteria that are at least equally important [188].

All in all, when explainees receive explanations, they go through the process of explanation evaluation, through which they determine whether the explanation is useful, understandable and satisfactory or not.

5.3 Taxonomy of human subject evaluation of explanation

This section presents a preliminary taxonomy of evaluation approaches resulting from a literature review conducted to investigate how an explanation can benefit users of decision-making support systems. Human-Computer Interaction (HCI) perspective focusing on evaluation with human subjects was taken into account through the literature review to construct this taxonomy. The outcome of this review provided a better understanding of the essential decisions required to design the experiments of this research. Furthermore, it synthesised a guideline to reuse of and inspired by suitable best practices for describing the design of the experiment in this chapter more structurally and precisely. The remainder of this section argues relevant dimensions of explanation evaluation with human subjects by grouping identified characteristics into task-related, participant-related, and study design-related dimensions, adapted from [189].

5.3.1 Task-related characteristics

According to [190], there needs to be a significant match between the choice of evaluation and the specificity of the claim being made. Therefore, the type of human subject evaluations should be categorised in one of the following ***levels of task*** abstraction:

2. Application-grounded evaluation; which involves conducting human experiments to evaluate the quality of an explanation in the context of its end-task within a real application. It is not an easy task for HCI [191] because there is an essential baseline of human-produced explanations that help other humans complete the task.
3. Human-grounded evaluation; is a more straightforward human experiment that can be completed with lay humans when one wishes to test more general notions of the quality of an explanation. It can mitigate the compensation of highly trained domain experts and allow for both a bigger subject pool and less expenses [192].

Several studies have proposed different ***types of task*** for participants in human experiments to evaluate the quality of explanations [190] [193] [194] [195]. The tasks can be classified based on the information provided to the participant, and the information inquired in return. This classification can be listed as follows:

- Verification task: Participants are provided with input, explanation, and output and asked for their satisfaction with the explanation.
- Forced choice tasks: Participants are asked to choose from multiple competing explanations, one that they find of higher quality.
- Forward simulation/prediction: Participants are presented with an explanation and an input and need to correctly simulate/predict the system's output.
- Causal simulation/prediction: As an extension of forward simulation, participants are presented with input and its corresponding output. They are asked to explain or justify their simulation/prediction.
- Counterfactual simulation/prediction: Participants are presented with an explanation, input and its corresponding output, and an alternative output. They are asked to predict how input needs to be changed to obtain the alternative output.
- System usage tasks: Participants are asked to use the system and its explanations for its primary purpose, e.g. a decision-making situation. The quality of the explanation is assessed in terms of decision quality.

In [193], the ***involvement of the participant*** in the evaluation of explanations is categorised as:

- Feed-back setting: In this setting, the quality of the explanations is determined through the participant's feedback on actual explanations.
- Feed-forward setting: In this setting, no explanations are provided to the participant. Instead, they are asked to provide a reasonable explanation serving as a benchmark for algorithmic explanations.

Key measurement concepts are considered in the literature as another aspect of the evaluation, and suitable metrics for each concept are proposed. While in [195], "goodness", "satisfaction" and "measuring users' mental model" are suggested as key measurement concepts, another study [196] discussed "usability" as a key concept to assess the quality of explanations. Likewise, "causability" is introduced in [197] as a new concept to measure the quality of explanations.

5.3.2 Participant-related characteristics

Participants of a human experiment should have the expertise comparable with the competence of the potential expected users of the original evaluated system. ***User expertise*** determines what kind of cognitive chunks they have, that is, how they organise individual elements of information in their mental model [198]. For explanation evaluation experiments, participants can be divided into two different categories, as follows [199]:

- AI Novices (novices); refer to end-users who use AI products in daily life but have no (or very little) expertise in AI systems. These include end-users of intelligent applications like personalised agents (e.g., home assistant devices), social media, and e-commerce websites.
- Data Experts (including domain experts); include data scientists and domain experts who use AI for analysis, decision-making, or research. These users might be experts in certain domain areas of AI or experts in general areas of data science. Still, in our study, we consider users in this category to generally lack expertise in the technical specifics of Semantic Web technologies.

The experiences of the desired participants play an important role in determining the recruiting method and number of participants. While novices can be recruited in large numbers via crowdsourcing, in contrast, it is hard to identify and recruit domain experts. According to [200], the user study task may depend on the level of participant foresight. In most cases, participants have equal knowledge about the experiment's context based on provided information for them. Such experiments are usually suitable for novices. However, some experiments need further information, such as external facts or relevant experiences, to complete the experiment. Such a setting may be more suitable for data experts; however, it requires more control on participants' knowledge [200].

Incentivisation of participants is another relevant dimension. Participants may take part in a study because of study-related incentives (e.g. curiosity, sympathy, or entertainment), personal-incentive (e.g. professional interest or a promise made), or altruistic reasons (e.g. to benefit science, society, or others) [201]. Nevertheless, most of the time, using a monetary incentive for participants is more effective in participant recruitment [202]. Incentivisation should be considered according to study length, task demand, and participant expertise [203].

5.3.3 Study design-related characteristics

The study design may follow a qualitative, quantitative, or hybrid study approach. These approaches mainly differ in the way of conducting an experiment and collecting data. Qualitative research gathers data that is free-form and non-numerical, such as diaries, open-ended questionnaires, interviews and observations that are not coded using a numerical system. On the other hand, quantitative research gathers data that can be coded in a numerical form. Quantitative research commonly applies interviews/questionnaires that consist mainly of closed questions or rating scales for data gathering [204] [205].

Another crucial decision to make is whether an experiment will compare different data for each participant (such as success rates for different product designs) or data from each participant to the other participants (such as success rates for different age groups). The first approach is commonly referred to as a within-subjects design, and the second is known as a between-subjects design. A within-subjects study does not require a large sample size and does not consider differences across groups. The downside of a within-subjects design is that one may need to worry about "carryover effects", where performance in one condition impacts performance in another condition. A carryover effect might result from practice (improving performance) or fatigue (decreasing performance). Therefore, Counterbalancing needs to be applied to prevent a possible carryover effect. A between-subjects study is used to compare results for different participants in a larger sample size. In another type of between-subjects design, participants are randomly assigned to groups and receive different treatments, such as different prototype designs for the same product. One advantage of a between-subjects design is eliminating carryover effects because any potential carryover effects would impact both groups equally. A mixed design should be considered if neither a between-subjects nor a within-subjects design meets the experiment needs. A mixed design contains a between-subjects factor, such as gender, and a within-subjects factor, such as three trials distributed over time. Mixed designs can be a compelling technique, and because they may eliminate the need for separate studies for each question that arises, they can also be time-saving and cost-effective [206].

5.4 Experiments

This section describes two main experiments undertaken in this study by human subjects and shows the findings of each experiment. The first experiment was conducted to evaluate the interpretability of the explanations generated through OWL Explanation API. The second experiment was conducted to evaluate the interpretability of the explanations generated through the XPDA service, presented in two different formats, textual and visual, across different use cases.

5.4.1 Evaluation for interpretability of the explanations generated through OWL Explanation API

As discussed in Section 4.4, we could “generate justifications for entailments” on personal data access in the XPDA service using OWL explanation API [170]. Although evidence in [170] proves these explanations are difficult to understand even by domain experts, we conducted a (primitive/initial) experimental study via a similar setup of the experiment to:

- verify the finding for the purpose of our service and its application,
- decide whether it needs to offer extra assistance to users trying to understand these explanations, or

- estimate where extra steps could be inserted into an explanation in order to make it easier to understand.

5.4.1.1 Hypothesis

The hypothesis of this experiment is concerned with the extent to which initial explanation for the reason of personal data access, generated through OWL Explanation API, is understandable for AI novices and is stated as:

Hypothesis H1: AI novices cannot understand why data has been accessed, as it is explained through OWL Explanation API.

5.4.1.2 Methodology

A user experiment was carried out targeting a group of data expert users to test the hypothesis mentioned above. This experiment was conducted at the human-grounded level using a causal simulation method. Participants were presented with a scenario, a corresponding data access decision and an explanation for the reason for this access. They were asked to express their understanding of the explanation through a Likert scale. Participants' self-expressions were verified in detail by following up with face-to-face sessions using retrospective "think-aloud protocol" [207].

5.4.1.3 Participants

The study comprised five volunteers who were postgraduate students from the School of Computer Science and Statistics at Trinity College Dublin. The participants of this experiment were considered as data experts according to categories of participant's expertise discussed in Section 5.3.2.

5.4.1.4 Procedure

This experimental study was structured as follows:

- **Self-expression of understanding:** to assess participants' understanding, they were presented with following items as illustrate in Table 5.1:
    - the first use case of the motivating scenario discussed in Section 4.2 as an input,

    - an access decision for a corresponding personal data in the use case as an output, and
    - the explanation of the reason for accessing the personal data generated through OWL Explanation API.

  The participants were asked to rate the extent to which the generated explanation is easy to understand through a six-point Likert scale: ‘Very easy’, ‘Easy’, ‘Neither easy nor difficult’, ‘Difficult’, ‘Very difficult’, ‘Impossible’.

- **Face-to-face sessions using retrospective “think-aloud protocol”:** Participants were asked to say whatever comes into their mind as they read the explanation. This might include what they were looking at, thinking about and feeling. The notes were taken during the session and confirmed /verified by the participant at the end of the session.

Table 5.1. List of information provided to participants of the experiment

<table>
<tr>
<td><b>Scenario</b></td>
<td><i>Bob is a 25 years old man living in Dublin. Last year, he visited the Royal Hospital for a comprehensive health screening.</i>
<i>The Royal Hospital collected his health data (Diseases, Treatments, Allergies, ...) and demographics (Name, Gender, Date of birth, ...) with his explicit consent.</i>
<i>After a while, Bob visited the Royal Hospital again due to feeling a pain in his chest and he was hospitalised for half a day.</i>
<i>Ruth and Mary, as nurses, were on duty that day at the Royal Hospital. Ruth was assigned to Bob for caring for him.</i></td>
</tr>
<tr>
<td><b>Data access</b></td>
<td>
<table>
<tr><th>Who</th><th>Which data</th><th>Whom</th></tr>
<tr><td>Ruth</td><td>Disease</td><td>Bob</td></tr>
</table>
</td>
</tr>
<tr>
<td><b>Explanation</b></td>
<td>AccessRequest(?ar), Consent(?con), DataOwner(?do), DataRequester(?dr), Diseases(?dis), Location(?loc), Nurse(?nur), Organisation(?org), Patient(?pt), accessRequestFor(?ar, ?dis), hasAccessRequest(?dr, ?ar), hasLocation(?do, ?loc), hasLocation(?dr, ?loc), hasPersonalData(?do, ?dis), hasRole(?do, ?pt), hasRole(?dr, ?nur), obtainConsent(?hcc, ?con), providerFor(?dr, ?do), signedConsent(?do, ?con), workAt(?dr, ?org) -> hasDecision(?ar, grant), isMatchWith(?ar, pr-1)

pt-1 Type Patient
org-1 obtainConsent con-1
dtR-1 hasLocation loc-1
loc-1 Type Location
nur-1 Type Nurse
dtO-1 Type DataOwner
dtO-1 hasPersonalData dis-1
dtO-1 hasRole pt-1
dtR-1 Type DataRequester
dtR-1 hasAccessRequest req-1
obtainConsent Domain Organisation
req-1 accessRequestFor dis-1
accessRequestFor Domain AccessRequest
dtR-1 hasRole nur-1
dtR-1 workAt org-1
dtO-1 signedConsent con-1
dtO-1 hasLocation loc-1
dtR-1 providerFor dtO-1
dis-1 Type Diseases
con-1 Type Consent</td>
</tr>
</table>

5.4.1.5 Results

Not surprisingly, the self-expression outcome showed that four out of the five responses from participants measured the explanation of being “Very difficult”. The other one measured it as “Impossible” to understand.

The qualitative synthesising of the notes from follow up sessions demonstrated that participants commonly but in different phrases declared their disability to understand the explanation.

This results can be considered as an immediate defeat for the interpretability of the explanation generated through the OWL explanation for the participants. Therefore, it can be concluded that:

If data experts face difficulty understanding the explanation generated through OWL explanation API, then it would be even further burdensome to novices to comprehend it. Furthermore, if the inference of the reason for data access is complicated for a data expert, so novices absolutely cannot do it.

Therefore, it shows that the hypothesis of this experiment is possible to accept.

5.4.2 Evaluation for interpretability of the explanations generated through the XPDA service

5.4.2.1 Introduction

As discussed earlier in this chapter, the interpretability of explanation can be defined as the degree to which a participant can understand the cause of personal data access by generated explanation through the XPDA service. Also, inspired by the way that humans assess the quality of explanation, discussed in Section 5.2, we determined the following key measure concepts and criteria for evaluation of interpretability:

- Usability; as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use” [208].
- Understandability; as “the degree to which a human can understand a decision made by a model” [209].
- Satisfaction; as “the degree to which users feel that they understand the system or process being explained to them” [195].

Therefore, an experiment was designed following a quantitative approach to determine whether the explanation generated through the XPDA service is usable, understandable and satisfactory or not. This experiment was conducted at a human-grounded level through a mixed design of the experiment to exploit all merits of both within-subject and between-subject approaches mentioned in Section 5.3.3.

5.4.2.2 Experiment Instruments

This section describes the instruments adopted in this experiment: Qualtrics [38], an online survey platform, was used to build the surveys/questionnaires of the experiment, distribute them, and analyse responses; Prolific [37] was deployed for managing participant recruiting and payment; SUS [39] and ASQ [40] were applied as standard measures to evaluate usability and satisfaction in this experiment, respectively.

5.4.2.2.1 Qualtrics

Qualtrics[10] [38] is an online survey platform that provides a researcher with the ability to create a survey for a distribution channel. It collects and compiles the responses from various participants across different channels, validates them, then updates the overall survey results and presents them to the researcher that created the survey.

5.4.2.2.2 Prolific

Prolific[11] [37] is an established platform to handle online subject recruitment and payment, which explicitly caters to researchers. The researchers can use any web-based platform to collect the actual data. It combines good recruitment standards across diverse participants with a reasonable cost at a fixed fee according to the average time taken to task completion. The time required for an experiment is initially estimated by the experimenter but is then updated with the actual time taken once participants make submissions. At the time of current research, the minimum payment per hour was 5 GBP or 6.50 USD, with fractions of hours requiring proportionally smaller payments.

[10] https://www.qualtrics.com

[11] http://www.prolific.co/

Researchers can filter the participants according to the specific sampling requirement of their experiment based on the pre-screening questions and/or their acceptance score and, reputation score was recorded and updated by Prolific. Also, researchers can reject a participant's submission if they can reasonably be justified in thinking that the participant has made little effort, failed multiple attention checks or has lied their way into the experiment.

Also, if a study has to be repeated, or different treatments are to be run sequentially, or if experiments are run using the same account, a screener allows for excluding subjects who participated in specified previous studies. All these facilities make Prolific an appealing tool for most HCI research [210] [211] [212] [213].

5.4.2.2.3 SUS

System Usability Scale (SUS) [39] is one of the most popular post-study standardised questionnaires to quickly and easily assess the usability of a given product or service by practitioners. The original SUS is composed of ten statements, shown in Table 5.2, that are scored on a 5-point Likert scale for strength of agreement. These statements cover various aspects of system usability, such as the need for support, training, and complexity. Thus, it has a high level of face validity for measuring a system's usability. The odd-numbered items have a positive tone, while the tone of the even-numbered items is negative. Participants should be asked to record their immediate response to each item rather than thinking about them for a long time. According to the SUS scoring method, all ten items should be responded to by participants. If participants couldn't respond to an item for some reason, the centre point of the scale should be selected.

Table 5.2. System Usability Scale questionnaire

Item	Item description
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

SUS yields a single number representing a composite measure of the overall usability of the system being studied, and scores for individual items are not meaningful on their own. Final scores for the SUS can range from 0 to 100, where higher scores indicate better usability. Because the statements alternate between the positive and negative, care must be taken when scoring the survey. The first step in scoring SUS is to determine each item's score contribution, which will range from 0 to 4. For positively worded items (odd numbers), the score contribution is the scale position minus 1. For negatively worded items (even numbers), the score contribution is five minus the scale position. To get the overall SUS score, the sum of the item score contributions should multiply by 2.5. Thus, the overall SUS scores range from 0 to 100 in the 2.5-point increments. The Overall SUS score ranging from 0 to 100 often leads researchers to interpret the SUS scores as percentages, which is not true. The percentage could consider using Percentile range [214].

With the advent of large sample data sets of the SUS scores, there have been a few attempts to provide a "grading scale" for their interpretation. A grading scale was developed in [215]. Also, the idea of using words instead of numbers to describe the SUS scores was proposed in the same research as an adjective scale. Later, both scales were revised according to a comprehensive investigation [216], and the outcome was new ranges for grading and adjective scale illustrated in Figure 5.1.

Another variation on using words to interpret the row SUS score considers the terms of "Acceptable" or "Not Acceptable" corresponds to roughly above 71 and below 51.7, respectively. The range between 51.7 – 71 is designated as "Marginal".

Another approach of interpreting the raw SUS scores is to convert them into percentile ranks. In [217] the large dataset of the SUS scores was taken and "normalised" to allow for percentile ranks. Percentile ranks tell how well a raw score is compared to others in the database.

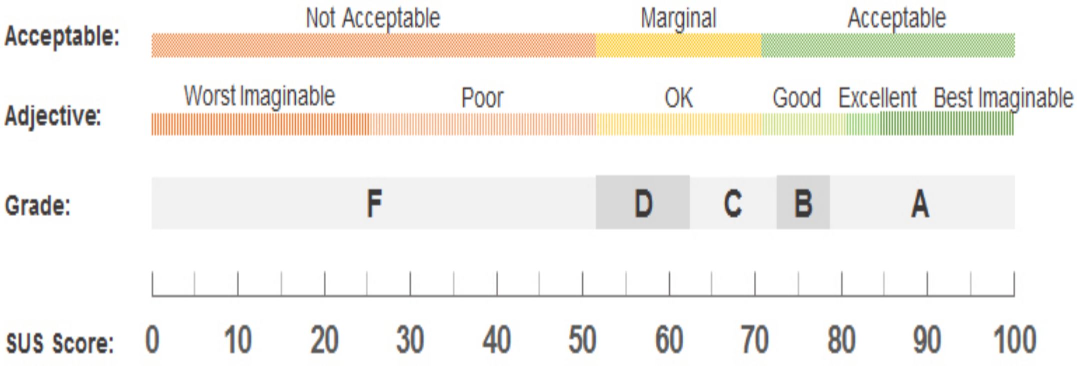


Figure 5.1. Different interpretations of the SUS score [218]

An early assessment (using coefficient alpha) of SUS indicated the reliability of 0.85 [218]. More recent estimates using larger samples have found its reliability to be just over 0. 9 [215] [216].

SUS has several attributes that make it a good choice for general usability practitioners:

1. It is technology agnostic and flexible enough to assess a wide range of technologies and services.
2. It is relatively quick and easy to use by both study participants and researchers.
3. It is non-proprietary and does not require any license fee.

5.4.2.2.4 ASQ

After-Scenario Questionnaire (ASQ) [40] assesses user satisfaction during participation in scenario-based usability studies. It comprises three statements that are scored on a 7-point Likert scale of the strength of agreement, covering various aspects of satisfaction such as overall ease of task completion, completion time, and support information, illustrated in Table 5.3.

Table 5.3. After-Scenario Questionnaire

Item	Item description
1	Overall, I am satisfied with the ease of completing the tasks in this scenario.
2	Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario.
3	Overall, I am satisfied with the support information (online help, messages, documentation) when completing the tasks.

The overall ASQ score is the average of the responses to three items which takes a value between 1 and 7, where lower scores indicate better satisfaction. ASQ has acceptable psychometric properties of reliability, sensitivity, and concurrent validity, with reliability measure in a range from 0.9 to 0.96 [219]. ASQ is non-proprietary and does not require any license fee, but anyone using it should cite and acknowledge the source of the measure.

5.4.2.3 Experiment Hypotheses

This section explains the primary hypothesis of the experiment and its derivative as secondary hypotheses.

5.4.2.3.1 Primary Hypothesis

The primary hypothesis of this experiment is concerned the extent to which AI novices can understand the cause of their personal data access presented as an explanation generated through XPDA, and is stated as:

Hypothesis H1: The explanation of the reason for personal data access generated through XPDA is interpretable for AI novices.

5.4.2.3.2 Secondary Hypotheses

Three secondary hypotheses were constructed to focus on the main evaluation criteria for interpretability; usefulness, understandability and satisfaction.

Hypothesis H1.1: Different forms of explanation of the reason for personal data access generated through the XPDA service are usable for AI novices, discussed in Section 5.4.2.8.

Hypothesis H1.2: AI novices can understand different forms of explanation of the reason for personal data access generated through the XPDA service, discussed in Section 5.4.2.9.

Hypothesis H1.3: AI novices are satisfied with different forms of explanation of the reason for personal data access generated through the XPDA service, discussed in Section 5.4.2.10.

5.4.2.4 Experiment Design

For the design of this experiment, use cases of the motivating scenario, described in Section 4.2, and presentation forms of their corresponding explanation are considered as independent variables/factors of the experiment. Hence, various use cases of the scenario and different presentation forms of generated explanation for reasons of data access in a sample access decision corresponding to each use case (textual or visual) are assumed as possible values of these independent variables/levels of the factors. Therefore, all possible combinations of these values composed a set of treatments for the experiment. The treatments are used to measure three criteria of the experiment (usability, understandability, and satisfaction), which are considered as dependent variables of the experiment. The results of the evaluation of these criteria are used to verify the hypotheses of the experiment exhaustively.

Two treatments are included in the design of each questionnaire of the experiment to increase the number of experiment units by recruiting the same number of human subjects. Extra attention is taken into account to avoid the learning effects of encountering neither the same scenario nor the same presentation form of the explanation in a questionnaire. Consequently, 12 distinct questionnaires with the same structure were designed. Details of the questionnaires' structure will be discussed in the next section.

The procedure of identifying valid treatments for this experiment can be formally described as follows:

If a set of use cases is defined as S with three different scenarios as its levels/values;

$$S = \{s_1, s_2, s_3\}$$

and set of presentation forms is defined as P, consists of textual presentation as t and visual presentation as v;

$$P = \{t, v\}$$

Then set of treatments, T, can be calculated using "*combination of* n *taken* r" in the set of possible scenarios and the set of possible forms of presentation.

$$C_X(n, r) = \frac{n!}{(n-r)!\,r!}$$

In the equation above, n represents the number of objects in the set X and r represents the number of objects taken at a time.

Thus:

$$C_S(3,1) = \frac{3!}{(3-1)!\,1!} = 3$$

$$C_P(2,1) = \frac{2!}{(2-1)!\,1!} = 2$$

The *multiplication principle of combinatorics* argues that if there are x ways of doing one thing and y ways of doing another, then the total number of ways of doing both things is $x * y$. So:

$$|T| = C_S(3,1) \, * \, C_P(2,1) = 3 * 2 = 6$$

Finally, a set of treatments, T, can be identified as:

$$T = \{\, s_1t \,,\, s_1v \,,\, s_2t \,,\, s_2v \,,\, s_3t \,,\, s_3v\}$$

Then, a possible arrangement of scenarios in each questionnaire is considered as *k*-permutations of *n*, which is defined as arrangements of a fixed length *k* of elements taken from a given set of size *n*:

$$P_x(n, k) = \frac{n!}{(n-k)!}$$

Therefore, to expose two mutually exclusive scenarios to the participants in each questionnaire, six different arrangements need to be considered:

$$P_S(3,2) = \frac{3!}{(3-2)!} = 6$$

Likewise, the arrangement of two various forms of explanation in a questionnaire is considered as *k*-permutations of *n*, too:

$$P_P(2,2) = \frac{2!}{(2-2)!} = 2$$

Finally, using the *multiplication principle* of combinatorics avoids the learning effect in each questionnaire. That means we need to multiply in permutations of scenarios and presentation forms to find the appropriate permutations of treatments, P_T:

$$|P_T| = P_S(3,2) * P_P(2,2) = 6 * 2 = 12$$

In summary, the user experiment was conducted using a set of 12 questionnaires with the same structure corresponding to the following arrangements to fulfil all requirements and verify all hypotheses.

$$P_T = \left\{ \begin{matrix} s_1t\ s_2v\ ,\ s_2v\ s_1t\ ,\ s_1t\ s_3v\ ,\ s_3v\ s_1t\ ,\ s_1v\ s_2t\ ,\ s_2t\ s_1v, \\ s_1v\ s_3t\ ,\ s_3t\ s_1v\ ,\ s_2t\ s_3v\ ,\ s_3v\ s_2t\ ,\ s_2v\ s_3t\ ,\ s_3t\ s_2v \end{matrix} \right\}$$

For example, when a participant engaged in the experiment and was asked to answer to questionnaire $\boldsymbol{s_1t\ s_2v}$, firstly, s/he was provided with the scenario of **use-case 1**, the corresponding explanation of the reason for an access decision in this use-case presented in the **textual** format and was asked to answer the related set of questions (will be discussed in detail in next section). Then, s/he was provided with the scenario of **use-case 2** along with the corresponding explanation of the reason for an access decision in this use case presented in the **visual** format and was asked to answer the related set of questions. Therefore, s/he encountered neither the same use case nor the same presentation form of the explanation in the questionnaire, and her/his answers were not influenced by learning effects.

5.4.2.5 Questionnaire Design

Using a questionnaire facilitates gathering the data from a large number of participants at a relatively low cost and efficient time. Qualtrics was used as an online platform to build a confidential online survey, distribute it and analyse responses in this experiment. As discussed in the previous section, 12 questionnaires were designed to conduct this experiment to cover different combinations of treatments and their permutations. The structure of these questionnaires includes the following blocks:

- **Participant Information Sheet**; provides potential participants with the necessary understanding of the motivation and procedures of the study and sources of information to answer any further questions to allow them to give informed consent.
- **Consent Form**; repeats the information mentioned above to ensure the key points are understood. Participants become aware of the reason for conducting the study, the procedures involved, potential risks, and how they can get more information about the study. Each participant must acknowledge this understanding by clicking a button through a web-form recorded electronically as informed consent. The content of the consent form can be found in Appendix B.
- **Pre-Study Questions**; aim to better understand specific characteristics of the participant. Several questions are developed to gather participants' demographic information such as age range and gender, their incentive and enthusiasm to understand the details of their personal data access, and their level of relevant competencies such as English reading skills, the latest degree of education and familiarity of Graph visualisation.
- **Task Description Sheet;** gives information or instructions to the participants on how to respond to the upcoming questions related to pre-defined treatments of the experiment.
- **Tasks Questions**; consist of questions built to measure three criteria of the experiment (usability, understandability and satisfaction) and resembled respectively according to the proportionate treatment of the experiment. Two series of task questions are provided based on the permutation of treatments consecutively. Details of constructing task questions per each criterion will be explained later in Sections 5.4.2.8 to 5.4.2.10.

The questionnaires were reviewed and approved by "The School of Computer Science and Statistics Research Ethics Committee" of Trinity College Dublin.

5.4.2.6 Participant Recruiting

The participants' recruitment for this experiment was crowdsourced by targeting AI novices using Prolific, described in Section 5.4.2.2.2. The age of the participants was restricted to be +18 due to research ethics advice. No other restriction was put in place for the pre-screening of participants to increase inclusion and diversity.

The estimated participation time was granted 20 minutes, with 67 minutes maximum participation time allowed accordingly by Prolific. A monetary incentive of GBP 2.5 was paid per participation. This payment was appraised as a "good" level of reward per hour by Prolific[12] at the time of experiment design.

The abovementioned procedure of recruiting was replicated separately per each one of 12 distinct questionnaires of the experiment. To ensure participants' uniqueness, once a person took part in a questionnaire of this experiment, s/he was restricted to participate in the rest of the questionnaires through custom pre-screening.

Therefore, 60 participants were recruited initially to carry out this experiment. Two submissions were rejected due to a quite fast completion time (less than 3 minutes) than the average completion time per participation (18 minutes and 37 seconds). Then, two new participants replaced them. According to participants' self-expression among pre-study questions of the experiment, the diversity of participants regarding their gender was 42% males, 57% females and 1% others. Further specifications of participants can be categorised in corresponding groups as listed in the following three tables:

[12] www.prolific.co [accessed at 20.7.2019]

Table 5.4. Participants distribution based on their age range

Age range	18 - 24	25 - 34	35 - 44	45 - 54	55 - 64	65 and older
Percentage of participants	52%	23%	15%	5%	2%	3%

Table 5.5. Participants distribution based on their education degree

Education degree	Less than high school	High school graduate	Some college	2-year degree	4-year degree	Professional degree	Postgraduate/Master	Doctorate
Percentage of participants	0%	22%	25%	8%	18%	12%	0%	15%

Table 5.6. Participants distribution based on their English reading skill

English reading skill	Basic	Fluent	Professional	Native
Percentage of participants	3%	30%	27%	40%

5.4.2.7 Pilot Study

A pilot study is best used to identify the smaller issues that can be addressed reasonably quickly before the actual research begins. After a questionnaire is developed, it is imperative to do a pilot study (also known as pre-testing the survey) to ensure that the questions are clear and unambiguous. There are two different areas of interest within a pilot study: the questions themselves and the interface of the survey [220].

A three-stage process of pre-testing a survey suggested by [221] is adopted by the pilot study for this experiment. This pilot study involved three volunteers, native English-speaking postgraduate students from the School of Computer Science and

Statistics at Trinity College Dublin, reviewing questionnaires, examining the questions' clarity and motivation, and evaluating the quality of both the survey tool and implementation procedures. Since most measures used to assess the criteria of this experiment are selected from the standard and well-established HCI approaches, just a few issues were discovered in this pilot study. The pilot study identified minor confusing or misleading issues in some of the questions, which were solved by rewording them. Also, the order of some items in measuring understandability is changed according to similar suggestions of pilot study participants.

5.4.2.8 Evaluation for the usability of the explanations

As discussed in Section 5.2, people judge explanations based on their usefulness [186], but there are a few research and practice on how to measure usefulness due to the lack of metrics to measure it [222]. However, researchers in HCI have recognised the close connection between usefulness and usability [223], and they consistently defined usefulness as an involving attribute of usability and actual systems use [224] [225]. According to the results of both the quantitative and qualitative analyses in [222], the concepts of usefulness and usability are closely linked. While some studies like [226] applied a self-expression approach to evaluate usefulness of explanations, the SUS score was used in some studies, such as [227] and [137], to evaluate usefulness and general usability of explanation for the intelligible services.

The following sections describe how the usability of different presentation forms of explanation generated through the XPDA service was evaluated using the SUS score.

5.4.2.8.1 Hypothesis

Hypothesis H1.1: Different forms of explanation of the reason for personal data access generated through the XPDA service are usable for AI novices.

5.4.2.8.2 Methodology

This evaluation was carried out targeting a group of AI novices to test the hypothesis mentioned above. It was conducted based on the System Usability Scale (SUS) by presenting the participants with a treatment and asking them to specify their level of agreement to any of the SUS statements. The hypothesis was tested on all possible treatments, and the results were compared using statistical tests to check the significance of the differences between them.

5.4.2.8.3 Procedure

A participant was alternately presented with two treatments, each of which was followed by a ten-item questionnaire based on SUS. Since the evaluation results might be affected by the order in which the questionnaire items were presented, items of the questionnaire were fully shuffled in a random order every time they were presented to the participants to have more systematic control on these types of effects.

The original SUS items refer to “system”, but substituting the word “website” or “product,” or using the actual website or product name were shown in different studies [214]. It is argued by [216] that proposing minor changes to the wording of the SUS items do not affect the resulting scores if any of these types of minor substitutions is kept consistent across the items. Therefore, all occurrences of the “system” were replaced with “explanation” consistently across all SUS questionnaire items to appropriately fit this evaluation. Table 5.7 illustrates the amended version of the SUS questionnaire items used in this evaluation.

Then, the participant was asked to specify her/his level of agreement to each item in five points: (1) Strongly disagree; (2) Disagree; (3) Neither agree nor disagree; (4) Agree; (5) Strongly agree.

Finally, the participant’s responses were scored based on the specific scoring method of SUS, narrated in detail in Section 5.4.2.2.3.

Table 5.7. Amended version of the System Usability Scale questionnaire used in this thesis

Item	Item description
1	I think that I would like to use such an explanation frequently.
2	I found the explanation unnecessarily complex.
3	I thought the explanation was easy to use.
4	I think that I would need the support of a technical person to be able to use this explanation.
5	I found the various parts of this explanation were well integrated.
6	I thought there was too much inconsistency in this explanation.
7	I would imagine that most people would learn this explanation very quickly.
8	I found the explanation very cumbersome to use.
9	I felt very confident using the explanation.
10	I needed to learn a lot of things before I could get going with this explanation.

5.4.2.8.4 Results

After collecting all responses from participants related to the usability of explanation generated through the XPDA service, the SUS scores were calculated and analysed. Table 5.8 depicts the descriptive statistics of data and shows that we can be 95% confident that overall mean of the SUS scores for the true population (AI novices) is within a range of 78.75 ± 2.52.

Table 5.8. Descriptive statistics of SUS scores per treatment

Use case	Presentation	Count	Mean	Standard Deviation	Minimum	Maximum	Confidence Level (95.0%)
Use-case 1	Textual	20	81.75	14.38	42.50	100.00	6.73
	Visual	20	77.63	13.80	50.00	97.50	6.46
Use-case 2	Textual	20	80.38	10.80	62.50	100.00	5.05
	Visual	20	74.00	15.84	45.00	95.00	7.41
Use-case 3	Textual	20	82.88	12.59	50.00	100.00	5.89
	Visual	20	75.90	15.14	28	92.5	7.09
Overall	Overall	120	78.75	13.93	28	100	2.52

The first and foremost outcome of analysis (Table 5.9) presents that in all treatments of this experiment and consequently overall, the percentages of above-average scores are higher than below-average ones. It shows the SUS scores given by the majority of participants are higher than the average of the SUS scores in each treatment.

Table 5.9. Percentage of above and below average SUS scores per treatment

Use case	Presentation	Below Average	Above Average
Use-case 1	Textual	32%	68%
	Visual	43%	57%
Use-case 2	Textual	44%	56%
	Visual	44%	56%
Use-case 3	Textual	32%	68%
	Visual	35%	65%
Overall	Overall	34%	66%

The scores are analysed for detailed interpretation based on different rankings scales, discussed in Section 5.4.2.2.3. As illustrated in Figure 5.2, 74% of participants found the usability of explanation "Acceptable" and 20% of participant ranked it as "Marginal", while 6% of participants found it "Not Acceptable".

The similar result came out through adjective ranking, where 6% of participants considered the usability of explanation generated through the XPDA service as "poor" while 20% of participant measured it as "Ok" and 74% of them scored above good as 19%, 13% and 43% of responses are ranked as "Good", "Excellent", "Best Imaginable", respectively. Meanwhile, 66% of participants scores are graded as 'A' following with 8% as 'B' and 9%, 12% and 6% as 'C', 'D' and 'F', respectively.

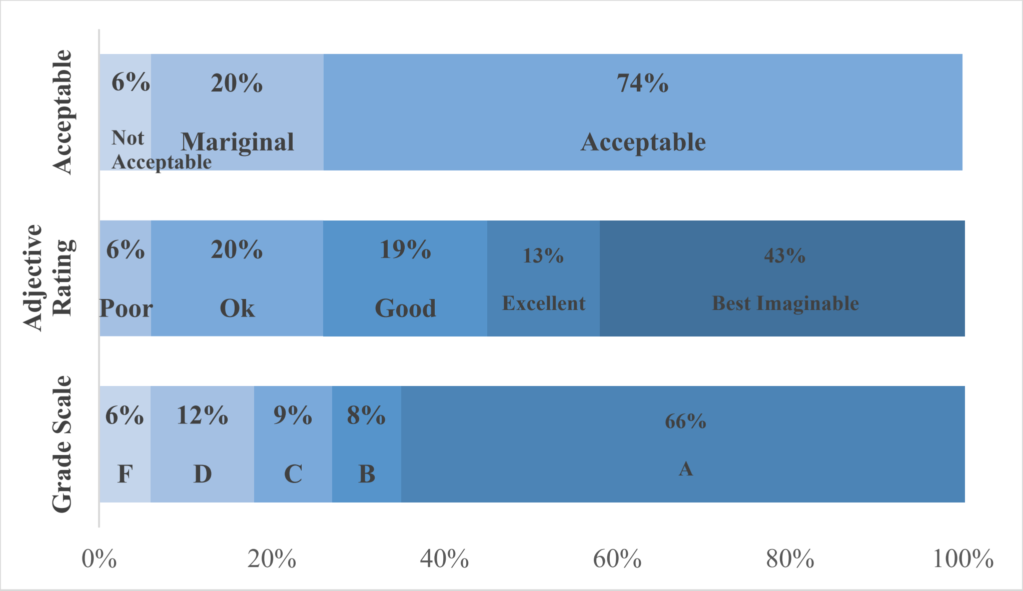


Figure 5.2. Percentage of overall SUS scores in different scales

Although, all these results confidently confirm the Hypothesis H1.1, further analysis of usability scores across different treatments was investigated as follows:

- Analysing corresponding results for various treatments of this evaluation (Table 5.10 – 5.12) depicts the trend mentioned above of usability scores remain similar across all different rankings scales, and the majority of participants strongly agreed that the explanation of the reason for data access corresponding to all scenarios are usable for them either if they are presented in the textual or visual format.
- One-way Analysis of Variances (ANOVA) was conducted to compare usability scores between explanations corresponding to different use cases when they are represented in a similar format. The degree of freedom between the groups was 2 and within groups was 57. The results indicate that there is no significant difference between usability of explanations across various use cases no matter whether they are presented in textual format (*P-value = 0.82*) or visual format (*P-value = 0.75*).
- Another level of analysis was conducted to investigate the effect of presentation form of explanation on usability scores per each use case through several independent-samples t-tests. The results (Table 5.13) depict that there is no significant effect for the presentation form of the explanations on usability scores in any use cases.

Table 5.10. Percentage of acceptance rate for SUS scores per treatment

Use case	Presentation	Not Acceptable	Marginal	Acceptable
Use-case 1	Textual	5%	10%	85%
	Visual	10%	15%	75%
Use-case 2	Textual	0%	30%	70%
	Visual	10%	30%	60%
Use-case 3	Textual	5%	10%	85%
	Visual	5%	25%	70%
Overall	Overall	6%	20%	74%

Table 5.11. Percentage of Grad scale for SUS scores per treatment

Use case	Presentation	F	D	C	B	A
Use-case 1	Textual	5%	10%	0%	5%	80%
	Visual	10%	10%	5%	20%	55%
Use-case 2	Textual	0%	10%	20%	10%	60%
	Visual	10%	20%	15%	0%	55%
Use-case 3	Textual	5%	10%	0%	0%	85%
	Visual	5%	10%	15%	10%	60%
Overall	Overall	6%	12%	9%	8%	66%

Table 5.12. Percentage of Adjective scale for SUS scores per treatment

Use case	**Presentation**	Worst Imaginable	Poor	Ok	Good	Excellent	Best Imaginable
Use-case 1	Textual	0%	5%	10%	15%	15%	55%
	Visual	0%	10%	15%	30%	15%	30%
Use-case 2	Textual	0%	0%	30%	15%	5%	50%
	Visual	0%	10%	30%	15%	10%	35%
Use-case 3	Textual	0%	5%	10%	15%	15%	55%
	Visual	0%	5%	25%	25%	15%	30%
Overall	Overall	0%	6%	20%	19%	13%	43%

Table 5.13. T-Test results for SUS scores of different presentation forms of each use case

Use case	Presentation	T values	P-values
Use-case 1	Textual	t (38) = 0.93	0.18
	Visual		
Use-case 2	Textual	t (38) = 1.49	0.07
	Visual		
Use-case 3	Textual	t (38) = 1.58	0.06
	Visual		

5.4.2.9 Evaluation for understandability of the explanations

The following sections will describe how to assess the extent to which participants understand different forms of explanation generated through the XPDA service.

5.4.2.9.1 Hypothesis

Hypothesis H1.2: AI novices can understand different forms of explanation of the reason for personal data access generated through the XPDA service.

5.4.2.9.2 Methodology

To test the aforementioned hypothesis, this evaluation was conducted targeting a group of AI novices. It was performed by presenting the participants with a treatment and asking them to answer three questions per treatment. The hypothesis was tested on all possible treatments, and the results were compared using statistical tests to check the significance of the differences between them.

5.4.2.9.3 Procedure

A participant was alternately presented with two treatments, each of which was followed by three questions which were proposed as follows:

- Question 1; designed as a 4-option Multiple-Choice Question Type A [228] to assess the extent to which the participants understood explicitly mentioned cognitive chunks of the explanation applying a causal simulation method in each treatment [190]. MCQ Type A is the most commonly used MCQ Type in which only the most appropriate option serves as the correct choice (the key).

- Question 2; designed as a 4-option Multiple-Choice Question Type X [228] for applying a forward simulation method in each treatment to assess the extent to which the participants understood the different level of compositionality of mentioned cognitive chunks for the explanation [190]. In MCQ Type X, which is known as multiple responses/answers MCQ, there may be more correct choices in a question instead of a single best choice.

- Question 3; designed to assess the extent to which the participants understood implicit cognitive chunks of the explanation applying a forward simulation method in each treatment [190]. This question is also proposed as a 4-option MCQ Type A.

An attention check question with an obvious correct response was added to the group of questions in each treatment to ensure scale validity by allowing to identify careless respondents and to screen them out prior to data analysis.

The study results might be affected by the order in which the items of the questionnaire are presented. So, the questionnaire items were fully shuffled in a random order every time they were presented to the participants to have more systematic control on these types of effects. Then, the participant was asked to choose the option(s) that seems more correct and appropriate for them across multiple choices.

Participants' responses for each question were scored using conventional methods in which correct responses were awarded a value of 1, whereas incorrect and omitted responses were awarded zero (00) value. To avoid guessing of answers by choosing all or none of the options in MCQ Type X, a suggested approach in [229] was followed in which if all or no options were selected, then no scores were awarded.

Mathematically, assuming an MCQ $question$, with n options, a participant's choice per each $option$ as i and correct choice per each $option$ as k; the score for the m^{th} $option$, $f_m(option)$, can be calculate as follows:

$$f_m(option) = \begin{cases} 1, & i = k \\ 0, & Otherwise \end{cases}$$

Consequently, $f(question)$, as the total score for the $question$, can be calculated as:

$$f(question) = \begin{cases} 0\,, & \forall\, m \leq n \mid f_m(option) = 0 \\ 0\,, & \forall\, m \leq n \mid f_m(option) = 1 \\ \sum_{m=1}^{n} f_m(option) \big/ n\,, & Otherwise \end{cases}$$

Finally, the score for measuring understandability of the explanation through this task, $f(task)$, can be calculated as the summation of the total score of each question, $f(question)$, which is ranging from 0 to 3, therefore:

$$f(task) = \sum_{q=1}^{3} f_q(question)\,, f(task) \,\epsilon\, [0,3]$$

In this evaluation, a threshold for sufficient understandability is defined as the minimum value of 2.0 to have solid and confident evidence. This threshold can be met if a participant answered at least 2 out of 3 questions correctly.

5.4.2.9.4 Results

The participants' responses to the questions were collected, and the scores were calculated based on the method described in the previous section. Descriptive statistics of data for all treatments of the task are depicted in Table 5.14.

Table 5.14. Descriptive statistics of understandability scores per treatment

Use case	Presentation	Count	Mean	Standard Deviation	Minimum	Maximum	Confidence Level (95.0%)
Use-case 1	Textual	20	2.41	0.47	1.75	3	0.22
	Visual	20	2.29	0.67	0.00	3	0.31
Use-case 2	Textual	20	2.11	0.85	0.75	3	0.40
	Visual	20	2.20	0.89	0.50	3	0.42
Use-case 3	Textual	20	2.41	0.56	1.50	3	0.26
	Visual	20	2.43	0.76	0.75	3	0.36
Overall	Overall	120	2.31	0.71	0	3	0.13

Analysis of the scores in Table 5.15 illustrates that in all treatments of this experiment and consequently overall, the absolute majority of the participants understood explanation generated through XPDA sufficiently. Although, these results are enough to confirm Hypothesis H1.2, further analysis of understandability scores across different treatments was investigated as follows:

Table 5.15. Percentage of understandability scores above and below value 2.0 per treatment

Use case	Presentation	Percentage of scores below value 2.0	percentage of scores at or above value 2.0
Use-case 1	Textual	30%	70%
	Visual	28%	72%
Use-case 2	Textual	38%	62%
	Visual	38%	62%
Use-case 3	Textual	28%	72%
	Visual	19%	81%
Overall	Overall	28%	72%

- One-way Analysis of Variances (ANOVA) was conducted to compare the understandability scores between explanations corresponding to different use cases when they are presented in a similar format. The degree of freedom between the groups was 2 and within groups was 57. The results indicate that there is no significant difference between understandability of explanations across various use cases no matter whether they are presented in textual format (*P-value = 0.25*) or visual format (*P-value = 0.66*). However, the reason for observed slight difference between results of Use-case 2 with results of other use cases can be the difficulty in the comprehension of the complex hierarchy in subsumption classes. It might also be due to the difference in participants' mental model of these hierarchies, which may cause unconscious bias to their previous knowledge, which might be varied with the data model in the use-case.
- Another level of analysis was conducted to investigate the effect of presentation form of explanation on understandability scores per each use case through several independent-samples t-tests. The results (Table 5.16) depict that there is no significant effect for the presentation form of explanations on understandability scores in any use cases.

Table 5.16. T-Test results for understandability scores of different presentation forms of each use case

Use case	Presentation	T values	P-values
Use-case 1	Textual	t (38) = 0.68	0.25
	Visual		
Use-case 2	Textual	t (38) = -0.32	0.38
	Visual		
Use-case 3	Textual	t (38) = -0.06	0.48
	Visual		

5.4.2.10 Evaluation for the satisfaction of the explanations

Explanation satisfaction is a contextualised, a posterior judgment of explanations representing the degree to which participants feel satisfied with the information being explained to them. Although the questions of ASQ, as a post-test questionnaire, ask participants in a way to measure their task-performance satisfaction, when users respond to post-test questionnaires, they tend to provide overall attitudes about the application in general and not necessarily their task performance [230]; hence their responses can be considered as perceived satisfaction. Therefore, post-test perceived satisfaction can show the extent to which participants are satisfied with the explanation.

5.4.2.10.1 Hypothesis

Hypothesis H1.3: AI novices are satisfied with different forms of explanation of the reason for personal data access generated through the XPDA service.

5.4.2.10.2 Methodology

To test the aforementioned hypothesis, this evaluation was carried out targeting a group of AI novices. It was conducted based on After-Scenario Questionnaire (ASQ) by presenting the participants with a treatment and asking them to specify their level of agreement to any of the ASQ statements. The hypothesis was tested on all possible treatments, and the results were compared using statistical tests to check the significance of the differences between them.

5.4.2.10.3 Procedure

A participant was alternately presented with two treatments, each of which followed by a three-item questionnaire based on ASQ. The study results might be affected by the order in which the items of the questionnaire were presented. So, the questionnaire items were fully shuffled in a random order every time they were presented to the participants to have more systematic control on these types of effects.

Then, the participants were asked to specify their level of agreement to each item in seven points: (1) Strongly agree; (2) Agree; (3) Somewhat agree; (4) Neither agree nor disagree; (5) Somewhat disagree; (6) Disagree; (7) Strongly disagree.

Finally, after the user has completed the questionnaire, the ASQ score was calculated using the average (arithmetic mean) of the three questions. Low scores are better than high scores due to the anchors used in the 7-point scales. If an item was skipped by participants, the ASQ score was calculated by averaging the score of remaining items.

In this evaluation, a similar approach of using words to interpret participants' satisfaction score took into account terms of "Satisfied" or "Unsatisfied" corresponding to roughly below four and above five, respectively. The range between 4 – 5 is designated as "Marginal".

5.4.2.10.4 Results

After collecting all responses from participants about their satisfaction of explanation generated through the XPDA service, ASQ scores were calculated and analysed. Table 5.17 shows the descriptive statistics of ASQ scores for participants' responses in different treatments:

Table 5.17. Descriptive statistics of ASQ scores per treatment

Use case	Presentation	Count	Mean	Standard Deviation	Minimum	Maximum	Confidence Level (95.0%)
Use-case 1	Textual	20	2.33	1.07	1	5	0.50
	Visual	20	2.13	1.07	1	5	0.50
Use-case 2	Textual	20	2.53	1.13	1	5	0.53
	Visual	20	2.43	0.85	1	4.67	0.40
Use-case 3	Textual	20	3.00	1.39	1	5.67	0.65
	Visual	20	2.32	1.12	1	5.33	0.52
Overall	Overall	120	2.46	1.12	1	5.67	0.20

For detailed interpretation, the scores are analysed based on the scale discussed above. As shown in Table 5.18, in all treatments of this experiment and consequently overall, the percentage of satisfied participants is higher than the percentage of unsatisfied participants. This ratio is 87% for satisfied participant versus 5% unsatisfied participant between all participant groups while the satisfaction of 8% is marginal.

Table 5.18. Satisfaction scales based on ASQ scores per treatment

Use case	Presentation	Unsatisfied	Marginal	Satisfied
Use-case 1	Textual	5%	10%	85%
	Visual	5%	0%	95%
Use-case 2	Textual	5%	15%	80%
	Visual	0%	5%	95%
Use-case 3	Textual	10%	15%	75%
	Visual	5%	5%	90%
Overall	Overall	5%	8%	87%

Table 5.19 shows the level of satisfaction of participants across all treatments of the experiment according to the average of their agreement level via ASQ. The table presents, in general, only 5% of participants exposed their disagreement about the satisfaction of explanations generated through XPDA, while the level of agreement for the rest of the participants about the satisfaction of explanations is calculated as 27%, 45% and 15% for “Strongly Agree”, “Agree” and “Somewhat Agree”. This trend of participants’ satisfaction remains similar across all different treatments.

Table 5.19. Satisfaction level of participants across all treatments

Use case	Presentation	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Use-case 1	Textual	0%	0%	5%	10%	5%	55%	25%
	Visual	0%	0%	5%	0%	20%	25%	50%
Use-case 2	Textual	0%	0%	5%	15%	10%	45%	25%
	Visual	0%	0%	0%	5%	20%	65%	10%
Use-case 3	Textual	0%	0%	10%	15%	25%	35%	15%
	Visual	0%	0%	5%	5%	10%	45%	35%
Overall	Overall	0%	0%	5%	8%	15%	45%	27%

- One-way Analysis of Variances (ANOVA) was conducted to compare satisfaction scores between explanations corresponding to different use cases when they are presented in a similar format. The degree of freedom between the groups was 2 and within groups was 57. The results indicate that there is no significant difference between the satisfaction of explanations across various use cases no matter whether they are presented in textual format (*P-value = 0.21*) or visual format (*P-value = 0.65)*.
- Another level of analysis was conducted to investigate the effect on the presentation form of explanation on satisfaction scores per each use case through several independent-samples t-tests. The results (Table 5.20) depict that there is no significant effect on the presentation form of explanations on satisfaction scores in any use cases.

Table 5.20. T-Test results for ASQ scores of different presentation forms of each use case

Use case	Presentation	T values	P-values
Use-case 1	Textual	t (38) = 0.59	0.28
	Visual		
Use-case 2	Textual	t (38) = 0.32	0.38
	Visual		
Use-case 3	Textual	t (38) = 1.71	0.05
	Visual		

5.5 Discussion

According to a systematic literature review of explanation in decision support systems [176], it is not typical to accompany an evaluation while proposing a new form of the explanation (lack of a proper evaluation in two-thirds of all analysed studies). It is not surprising because it is hard to determine and agree on the definition of correct or best explanation in most cases. The primary way to evaluate the provided explanations is to capture the subjective perception of users or to monitor the impact of the explanations in the user behaviour (which were predominant evaluation method occurring in more than half of the remaining analysed studies).

The "Right to Explanation" the reason of personal data access is a right to be given to any individual without considering his/her experiences and knowledge on AI/computer systems [25]. Therefore, both experiments were designed in a human-grounded level to test more general notions of the quality of explanations.

Since accuracy, consistency and completeness of the explanations generated through OWL Explanation API were assessed and confirmed in [170], the first experiment of this chapter assessed interpretability of the explanations generated through OWL Explanation API using a causal simulation method. The experiment confirmed that even domain experts with the background of computer science cannot understand this explanation with a high agreement between the participants about its complexity and difficulty. These results are aligned and consistent with [231], which showed that AI novice users prefer a more simplified explanation and representation interfaces.

The second experiment, which can be considered as a core experiment of this thesis, evaluated interpretability of the explanations generated through the XPDA service.

The experiment applied a mixed of within-subjects and between-subjects design to evaluate three well-accepted criteria of an interpretable explanation [32] [232] (usefulness, understandability and satisfaction) across all possible treatments corresponding to two presentation forms of the explanation and three pre-defined use cases.

According to the close connection between usefulness and usability, the usefulness of explanations was assessed through an amended version of the SUS questionnaire. Analysis of the results confirmed that most of the participants found the explanations generated through XPDA service sufficiently usable. This finding was consistent through all established ranking scale of SUS with similar trends across all treatments.

Assessing understandability of the explanations was more challenging due to the lack of standard metrics and even common method to evaluate it. Understandability of the explanations can depend on several latent parameters such as human cognitive function and can be subjective to the specific scenario/use case. Previous studies evaluated understandability of the explanations through self-expression of participants [233] [234], but it was shown that the participants tend to overestimate the depth of their understandings [235] as a case of a general overconfidence effect [236] [237]. People also seem to use misleading heuristics to assess how well they understand a system. Most notably, if they can see or easily visualise several components of a system, they are more convinced they know how it works [238]. An effective way of evaluating user understanding is to directly ask them about the decision-making process, which provides valuable information about their thought processes and mental models [195]. Therefore, several questions were designed to assess the understandability of the explanations generated through XPDA service. The results of this evaluation showed that most of the participants perceived the explanations with a high level of understanding.

As the last part of this experiment, participants' satisfaction with the explanation was evaluated through ASQ. The outcome confirmed that the participants were satisfied with the explanations generated through the XPDA service since they found them usable and understandable.

These evaluations confirmed all the hypotheses of the experiment and showed that the XPDA service could generate interpretable explanations of the reason for personal data access in both presentation forms (textual and visual) across different studied use cases.

5.6 Conclusion

This chapter presented the undertaken evaluations of the explanations generated in different phases of this thesis through two experiments with human subjects. The result of the first experiment highlighted the need to generate more interpretable explanation. The second experiment measured the interpretability of generated explanation through XPDA against most agreed criteria in the literature, namely usability, understandability, and satisfaction. A comprehensive user study was designed and conducted to evaluate these criteria across all possible treatments corresponding to two presentation forms of the explanation and three pre-defined use cases. The results of this study have shown that the explanations generated through XPDA service are sufficiently usable, perceived with a high level of understanding and satisfaction for the majority of participants. Therefore, all proposed hypotheses for the user study were confirmed, and RO4 of this thesis is fulfilled. However, there are several areas that future studies could explore further, which can be discussed as follows:

- As discussed in the previous section, both user studies are conducted to evaluate the quality of explanations in human-grounded level within the controlled environment of prototypical implementation of the motivating scenario. In future work, similar experiments need to be done in application-grounded level to evaluate the quality of explanations within a real application.
- There might be other methods to evaluate any of the selected evaluation criteria which can be investigated in future works. The comparison of several evaluation methods and their result would help to define a high-performance evaluation method to assess the quality of explanation not for TETs but also for other domains such as XAI. Likewise, although the selected criteria for evaluation are well accepted across different previous studies, but there are still several other key measurement concepts which need to be assessed for evaluating the quality of explanation.
- Although the main user study in this chapter is designed as a quantitative experiment, but expanding the study with post-experiment interviews, open-ended questions or more close-ended questions designed precisely by experts of different related domains as part of future work, can help to improve the validity of the outcomes of the studies.

6 Conclusion

This chapter concludes the thesis with a discussion on the extent to which the research question and objectives, discussed in Section 1.2, have been addressed through the presented work. The chapter also presents the resulting contributions, which were previously summarised in Section 1.4. It concludes with potential avenues for further work arising from the research presented within the thesis.

6.1 Fulfilment of Research Objectives

The research question guiding the work presented in this thesis, defined in Section 1.2 as:

"To what extent can a Semantic Web-based service enhance transparency to a human on her/his personal data access in an interpretable manner?"

Four research objectives were identified, which guided the work towards answering the research question. This section discusses the extent of their fulfilment based on work presented in previous chapters of the thesis.

6.1.1 Fulfilment of RO1

The first research objective (RO1) was to perform a literature review on the use of Semantic Web technologies to advance characteristics of access control and specifications of Transparency Enhancement Technologies for privacy-preserving purposes. This research objective was fulfilled by conducting two integrative reviews of previous studies on the above topics of interest. These reviews assessed and synthesised the literature to identify best practices, reveal the gaps and motivate the aim of this research and justify the research question and objectives, as described in Chapter 2.

The first review, discussed in Section 2.3, confirmed the competence of Semantic Web technologies in representing different access control models and in privacy policy specification and maintenance. This adequacy affirmed the motivation of exploiting Semantic Web technologies as a foundation and skeleton of the XPDA service. Also, comparing deployed approaches in reviewed literature identified the context-based approach and hybrid approach of using ontologies and rules as best practices for access control model and privacy specification, respectively. These approaches were used as essential methods in designing and implementing XPDA, as discussed in Chapter 3 and Chapter 4.

Meanwhile, the review acknowledged the necessity of providing data subjects with policy awareness and privacy implication to control the proper use of their data and preserve their privacy. The review discussed the need for improving the current access control approaches to provide data subjects with more transparent information on how and why their personal data was used. These findings led us to review the existing transparency enhancement technologies to improve our knowledge about their categorisation, representation and evaluation. This review, discussed in Section 2.4, featured that most of the existing TETs have been focused on privacy implication to help data subjects know what happens with their personal data. Meanwhile, we could not find any transparency enhancement approach in reviewed literature providing policy awareness that allows data subjects to know why their personal data disclosure and access. These outcomes justified RO2 and RO3 of this research as a need for design and implementation of a service, addressed in Chapter 3 and Chapter 4 respectively, to provide all above mentioned required aspects of transparency on personal data access in an interpretable manner. This part of the review also specified different metrics used to evaluate TETs; meanwhile, it demonstrated the need for a more comprehensive user-centric evaluation to measure the impact of these technologies. This outcome of review affirmed RO4 of this research which addressed in detail in Chapter 5.

6.1.2 Fulfilment of RO2

The architecture of XPDA was designed to extend data subject's control over their data access through enhancing transparency and discussed in Chapter 3. Motivated by literature review outcomes, Semantic Web technologies were used as a foundation of this architecture to provide sufficient flexibility to apply XPDA in different scenarios with few or no changes.

Meanwhile, putting human intelligence (such as domain experts and ontology engineers) in the architecture of XPDA augmented its scalability to deal with vast application domains and privacy policies. The strengths of different access control models and transparency enhancement technologies derived from their comparison in the literature review were considered as key attributes of XPDA. These fundamental attributes consist of:

- Exploiting context awareness for policy adaptation.

- The combined use of ontology constraints and rule-based approach for policy specification, enforcement and evaluation in the architecture of XPDA.

Also, this architecture fulfilled the lack of policy awareness and privacy implication of existing access control models by designing components for exposing the details of the access decision and explaining the justification of this decision as an ex-ante transparency enhancement approach in XPDA.

Consequently, the design of XPDA architecture satisfies RO2 of this research.

6.1.3 Fulfilment of RO3

The undertaken approaches to implement a prototype of the XPDA service on a motivating scenario in the health domain fulfilled RO3 of this study. Details of the implementation, discussed in Chapter 4, depicted how expected advantages of exploiting Semantic Web technologies in the design of XPDA was leveraged in practice to provide data subjects with more visibility on their personal data access. The brief list of these exploited competences of Semantic Web technologies in this prototype can be summarised as follows:

The determination of the concepts, their definitions and relationships comprising the vocabulary of data access control in the application domain is facilitated through ontological modelling. The graph presentation of the ontology made it more readable and understandable for privacy and legal regulatory experts to revise the model in the early stage of its development.

Specifying and representing the complex hierarchy of different entities of XPDA, such as personal data, and defining the complicated relations in contextual knowledge was simplified using OWL.

Integration of SWRL rule language and OWL constraints supported in the definition of privacy rules and specifying the policy.

Deploying Pellet reasoner, which supports OWL/Rule hybrid reasoning, enhanced the inference of latent knowledge about different concepts and their relations to provide more precise privacy enforcement and evaluation.

Querying the knowledge model using SPARQL retrieved the detailed information of access decision and justification for this access decision's entailment through OWL explanation workbench explained data access.

The explanation of the reason for a decision was provided in a human-readable format applying a verbalisation approach using labels in ontological knowledge model.

6.1.4 Fulfilment of RO4

A user study based on standard methodologies were conducted to evaluate the interpretability of outputs provided by XPDA and discussed in detail in Chapter 5 of this thesis.

The criteria for this evaluation and the best approach to measure them were determined by conducting another literature review focusing on the background of explanation evaluation in human behaviour, discussed in Section 5.2, and taxonomy of human subject evaluation of explanation, discussed in Section 5.3.

A comprehensive user study was designed to evaluate these criteria across all possible treatments corresponding to two presentation forms of the explanation and three pre-defined use cases, discussed in Section 5.4. While standard questionnaires were used to evaluate usability and satisfaction, a novel method was proposed to evaluate its understandability. Consequently, 12 different questionnaires were designed to conduct this user study. The experiments were conducted between and within 60 randomly recruited participants to increase the validity of the user study and the reliability of their results.

This study showed that the explanations generated through XPDA service are sufficiently usable, perceived with a high level of understanding and satisfiable for most participants. These experiments and their result fulfilled RO4 of this thesis.

6.2 Contributions

This section describes contributions from the research presented in this thesis, which were initially summarised in Section 1.4. The thesis yielded two contributions; design and development of the XPDA service, and design a user study to evaluate the interpretability of the explanation. The impact and extent of the contributions in terms of publications related to the work was listed in Section 1.4.3.

6.2.1 Design and development of the XPDA service

This thesis’s first and major contribution is to design and implement a service to provide data subjects with an interpretable explanation of their personal data access.

This service not only enhances the control of the data subject over their personal data access by leveraging Semantic Web technologies but also adopts the best practice of existing access control models by involving careful consideration on exploiting context awareness and policy specification. Moreover, this service extends the state of the art of TETs by offering a novel approach to provide detailed information about data access and explaining its justification which improves the visibility of the implication of privacy rules on access decisions. All of this information is represented in a way that could be perceived by non-expert users.

This service allows data subjects to benefit from their right to be informed about the collection and use of their personal data while they acquire their right to the protection of their personal data.

Although the primary beneficiaries of this service are data subjects, it can be exploited for auditing purposes by data controllers and service providers. Meanwhile, the research community can deploy and advance it in other domains.

6.2.2 Design a user study to evaluate the interpretability of the explanation

In this research, a comprehensive user study was designed to evaluate the extent to which participants can perceive the practical advantage of explanation generated through XPDA. The experiment of this user study deployed a quantitative approach to evaluate three well-agreed concepts of measurement for assessing the interpretability of generated explanation through XPDA. These criteria were identified as the results of an extensive review of different studies across several domains, including emerging XAI. The approaches for evaluation of usability and satisfiability were designed based on standard questionnaires. However, due to the lack of a standard experimental design method for evaluating the understandability, a novel method was adopted from synthesising various approaches of human cognition evaluation in social sciences. Different aspects of understanding, including explicit, implicit and compositional cognitive chunks of the explanation, were considered in the questionnaire design for this novel method.

The detailed description of all experimental design steps, including establishing treatments, calculating scores per each criterion, and analysing the results, amplified the flexibility and scalability of setting this user study up to evaluate the interpretability of explanations generated in other researches. The statistical analysis of results also applied to more investigation on the impact of different factors of the evaluation.

6.3 Limitations

This thesis describes several contributions and has the potential to yield a significant impact. However, there are some limitations in different phases of this research that should be noted, particularly with regard to guiding future work.

Deploying the XPDA service to control data access in any system/application relies on developing an ontological knowledge model of that system and is performed by an ontology engineer/expert, as discussed in Section 4.2. Therefore, the performance of XPDA highly depends on the quality of the developed ontology and consequently on the experience and domain knowledge of the ontology engineer/expert.

Another essential limitation identified during the implementation of XPDA, discussed in Chapter 4, was the lack of standard, general or even well-agreed methods/approaches that restricts the application of Semantic Web technologies. The immediate examples of this limitation can be seen in the:

- Lack of agreed and standard vocabularies to describe the key characteristics and categorisation of personal data and their processing categories in order to facilitate ontological modelling of services, like XPDA, and their interoperability in different use cases to demonstrate their effectiveness in various business settings. This gap is identified during the XPDA prototype implementation, where ontological knowledge modelling was developed in KR.
- Lack of a more standard and precise approach for verbalising the naming conventions used in ontological modelling, which depends on the linguistic features of the names used for individuals, classes, and properties in the input ontology. Interpreting machine-readable explanation, generated through OWL Explanation API, to human-readable presentation in natural language revealed this gap during AX implementation.

Further investigation is required to define these standards to extend the adaptivity of Semantic Web technologies for privacy-preserving purposes.

Moreover, the functionalities of some components were implemented manually in the prototype of XPDA. Although in some cases, it is carried out intentionally to simplify the procedure (such as the implementation of the Context Manager), in other cases, it occurred due to the lack of an automated method for the corresponding function creation. For example, identifying disparate data access policies issued by various stakeholders and representing them in rule languages, such as SWRL, has been carried out manually in the XPDA prototype and similarly in other application of Semantic Web technologies (based on best of our knowledge). This process would take a significant amount of time and labour in large scale applications. Further research to integrate existing approaches or propose a new approach for these functions can be considered as another potential future work.

Also, in implementing the XPDA prototype, the OWL Explanation API is deployed for generating machine-readable explanations for data access. Then these machine-readable explanations are transformed into human-readable explanations. Therefore, the limitation of OWL explanation API, such as dealing with multiple justifications [170] and reasoner benchmarking, has been inherited, affecting the quality of explanation generated through XPDA.

Finally, the prototype of XPDA is implemented based on a sample motivating scenario, and its experimental user study is also conducted in the controlled environment of this prototypical implementation. Therefore, while the research has shown that XPDA has clear promise, and the experimental results are positive, the effectiveness and efficiency of XPDA have not been tested in real applications across various domains. The potential future work can address exploiting XPDA in a real-world application to evaluate its efficiency with regards to different criteria, including fault tolerance, load balancing, and resource consumption.

6.4 Opportunities for future work

As discussed in the previous section, some gaps and limitations were realised during the research, which need to be further investigated in the future. Meanwhile, there are several opportunities identified to extend the capabilities of XPDA. Although the potential extent of each phase of this study was discussed separately in their corresponding chapters of the thesis, they can also be categorised as follows:

6.4.1 Future work to extend the design and development of the XPDA service

This research confirmed the positive impact of ex-post transparency on improving data subject control on their data access. The delivery mode of transparency in the prototype implementation of XPDA, discussed in Chapter 4, is applied on a demand-basis (Pull Mode). Improving the implementation to include a Push Mode, where data subjects can be notified of any event on their data, can enhance the level of transparency and interactivity of the XPDA service.

Another advancement of the XPDA architecture could offer data subjects more control over their data access, allowing them to change their privacy preferences as they comprehend the current implications. This improvement could help to enable the right to be forgotten.

Similarly, another potential expansion of the XPDA service can be offering ex-ante transparency, which provides data subjects with the anticipated consequences before disclosing data for intended data collection and processing.

6.4.2 Future work to extend the evaluation of the interpretability of XPDA

As discussed in chapter 5, conducted user study addressed the evaluation of explanations' quality in human-grounded level within a controlled environment of prototypical implementation of XPDA. Similar experiments need to be done in future works in application-grounded level within a real application requiring more complex and diverse privacy rules.

Deploying other alternative methods to evaluate the selected evaluation criteria and comparing their strengths, drawbacks, and results would help to define a high-performance evaluation method to assess the quality of explanation not only for TETs but also for other domains such as XAI. Also, several other key measurement concepts can be considered for evaluating the quality of the explanation.

Moreover, expanding the quantitative approach conducted in this thesis with post-experiment interviews, open-ended questions, or more close-ended questions designed precisely by experts of different related domains can improve the validity of the user study outcomes. This option can be considered as an outreach of current user study in future works.

6.5 Final remarks

It is the main hope of the author of this thesis that the proposed service can empower people's ability to control their data access by revealing more transparent details about these accesses. This will not be possible if it is not integrated with different commercial product and services. Therefore, the author's ambition is that the commercial providers of online products and services can realise the impact of this service on mitigating the privacy violation risk of their products and setting up the trust for their users and being willing to integrate XPDA with their products and services.

It is also hoped that the contributions of this study would benefit the research community, and researchers can employ the findings of this thesis in their research and apply their expertise to contribute to improving this study in the suggested future research directions.

References

[1] "Three challenges for the web, according to its inventor," *World Wide Web Foundation*, Mar. 2017. https://webfoundation.org/2017/03/web-turns-28-letter/.

[2] F. Brunton and H. Nissenbaum, *Obfuscation: A user's guide for privacy and protest*. Mit Press, 2015.

[3] J. Van den Hoven, M. Blaauw, W. Pieters, and M. Warnier, "Privacy and Information Technology," in *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, 2014.

[4] R. Bandara, M. Fernando, and S. Akter, "Privacy concerns in E-commerce: A taxonomy and a future research agenda," *Electron Markets*, Nov. 2019, doi: 10.1007/s12525-019-00375-6.

[5] M. Teltzrow and A. Kobsa, "Impacts of User Privacy Preferences on Personalized Systems," in *Designing Personalized User Experiences in eCommerce*, vol. 5, C.-M. Karat, J. O. Blom, and J. Karat, Eds. Dordrecht: Springer Netherlands, 2004, pp. 315–332.

[6] E. Toch, Y. Wang, and L. F. Cranor, "Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems," *User Model User-Adap Inter*, vol. 22, no. 1–2, pp. 203–220, Apr. 2012, doi: 10.1007/s11257-011-9110-z.

[7] danah boyd and K. Crawford, "Critical Questions for Big Data," *Information, Communication & Society*, vol. 15, no. 5, pp. 662–679, Jun. 2012, doi: 10.1080/1369118X.2012.678878.

[8] P. F. Wu, J. Vitak, and M. T. Zimmer, "A contextual approach to information privacy research," *Journal of the Association for Information Science and Technology*, vol. 71, no. 4, pp. 485–490, Apr. 2020, doi: 10.1002/asi.24232.

[9] J. Lane, V. Stodden, S. Bender, and H. Nissenbaum, *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. Cambridge University Press, 2014.

[10] M. Zimmer, "Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity," *Social Media + Society*, vol. 4, no. 2, p. 2056305118768300, Apr. 2018, doi: 10.1177/2056305118768300.

[11] D. J. Weitzner *et al.*, "Transparent accountable data mining: New strategies for privacy protection," 2006.

[12] S. Sackmann, J. Strüker, and R. Accorsi, "Personalization in privacy-aware highly dynamic systems," *Communications of the ACM*, vol. 49, no. 9, pp. 32–38, 2006.

[13] Web Foundation, "Personal Data: An overview of low and middle-income countries." 2017, [Online]. Available: http://webfoundation.org/docs/2017/07/PersonalData_Report_WF.pdf.

[14] C. Lazaro and D. L. Metayer, "Control over Personal Data: True Remedy or Fairy Tale," *SCRIPTed*, vol. 12, p. 3, 2015.

[15] B. Zhang, N. Wang, and H. Jin, "Privacy Concerns in Online Recommender Systems: Influences of Control and User Data Input," 2014, pp. 159–173, [Online]. Available: https://www.usenix.org/conference/soups2014/proceedings/presentation/zhang.

[16] L. Brandimarte, A. Acquisti, and G. Loewenstein, "Misplaced Confidences: Privacy and the Control Paradox," *Social Psychological and Personality Science*, vol. 4, no. 3, pp. 340–347, May 2013, doi: 10.1177/1948550612455931.

[17] T. W. Chen and S. S. Sundar, "This App Would Like to Use Your Current Location to Better Serve You: Importance of User Assent and System

Transparency in Personalized Mobile Services," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC, Canada, Apr. 2018, pp. 1–13, doi: 10.1145/3173574.3174111.

[18] M. D. Birnhack, "A Quest for A Theory Of Privacy: Context And Control," *Jurimetrics*, vol. 51, no. 4, pp. 447–479, 2011.

[19] G. Blank, W. H. Dutton, and J. Lefkowitz, "Perceived Threats to Privacy Online: The Internet in Britain, the Oxford Internet Survey, 2019," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3522106, Sep. 2019. doi: 10.2139/ssrn.3522106.

[20] B. Bellamy and C. Alonso, "Reframing data transparency." Centre for Information Policy Leadership and Telefónica Senior Roundtable, 2016, [Online]. Available: https://www.telefonica.com/documents/341171/2445513/CIPL+and+Telefonica+-+Reframing+Data+Transparency.pdf/9c007899-451c-4a5b-854d-784082e37bf7.

[21] M. Hansen, "Top 10 Mistakes in System Design from a Privacy Perspective and Privacy Protection Goals," in *Privacy and Identity Management for Life*, Berlin, Heidelberg, 2011, pp. 14–31, doi: 10.1007/978-3-642-31668-5_2.

[22] S. Fischer-Hübner, J. Angulo, F. Karegar, and T. Pulls, "Transparency, Privacy and Trust – Technology for Tracking and Controlling My Data Disclosures: Does This Work?," in *Trust Management X*, Cham, 2016, pp. 3–14, doi: 10.1007/978-3-319-41354-9_1.

[23] C. Andersson *et al.*, "Trust in PRIME," in *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology, 2005.*, Dec. 2005, pp. 552–559, doi: 10.1109/ISSPIT.2005.1577157.

[24] P. Murmann and S. Fischer-Hübner, "Tools for Achieving Usable Ex Post Transparency: A Survey," *IEEE Access*, vol. 5, pp. 22965–22991, 2017, doi: 10.1109/ACCESS.2017.2765539.

[25] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," *Official Journal of the European Union, vol. L119*, May 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC.

[26] G. Michener and K. Bersch, "Conceptualizing the quality of transparency," 2011.

[27] S. P. Wall, "Public Justification and the Transparency Argument," *The Philosophical Quarterly (1950-)*, vol. 46, no. 185, pp. 501–507, 1996, doi: 10.2307/2956360.

[28] R. Tagiuri, N. Kogan, and J. S. Bruner, "The Transparency of Interpersonal Choice," *Sociometry*, vol. 18, no. 4, pp. 368–379, 1955, doi: 10.2307/2785873.

[29] O. Svenson, "Process descriptions of decision making," *Organizational Behavior and Human Performance*, vol. 23, no. 1, pp. 86–112, Feb. 1979, doi: 10.1016/0030-5073(79)90048-5.

[30] M. Turilli and L. Floridi, "The ethics of information transparency," *Ethics and Information Technology*, vol. 11, no. 2, pp. 105–112, 2009.

[31] "Definition of Interpretability." https://www.merriamwebster.com/dictionary-/interpretability (accessed Jun. 17, 2020).

[32] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, 2019.

[33] B. Parsia, A. Fokoue, P. Haase, R. Hoekstra, and U. Sattler, *OWL 2 web ontology language structural specification and functional-style syntax*. W3C, W3C Recommendation, Dec, 2012.

[34] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean, "SWRL: A Semantic Web Rule Language Combining OWL and RuleML," *W3C Member Submission*, May 2004. https://www.w3.org/Submission/SWRL/.

[35] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, "Pellet: A practical owl-dl reasoner," *Journal of Web Semantics*, vol. 5, no. 2, pp. 51–53, 2007.

[36] M. Horridge, B. Parsia, and U. Sattler, "The OWL Explanation Workbench: A toolkit for working with justifications for entailments in OWL ontologies," p. 12, 2009.

[37] S. Palan and C. Schitter, "Prolific.ac—A subject pool for online experiments," *Journal of Behavioral and Experimental Finance*, vol. 17, pp. 22–27, Mar. 2018, doi: 10.1016/j.jbef.2017.12.004.

[38] I. Qualtrics, *Qualtrics*. Provo, UT, USA, 2013.

[39] J. Brooke, "SUS - A quick and dirty usability scale," *Usability evaluation in industry*, 1996.

[40] J. Lewis, "Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ," *SIGCHI Bull.*, vol. 23, pp. 78–81, Jan. 1991, doi: 10.1145/122672.122692.

[41] R. Gachpaz Hamed, K. Fatema, O. Conlan, and D. O'Sullivan, "Semantic Reasoning for Privacy-Preserving Personalisation," *11th International IFIP Summer School on Privacy and Identity Management*, p. 9, 2016.

[42] R. Gachpaz Hamed, H. J. Pandit, D. O'Sullivan, and O. Conlan, "Explaining Disclosure Decisions Over Personal Data," *18th International Semantic Web Conference*, p. 4, 2019.

[43] H. J. Pandit *et al.*, "Creating a vocabulary for data privacy: the first-year report of data privacy vocabularies and controls community group (DPVCG)," presented at the ODBASE 2019: The 18th International Conference on Ontologies, DataBases, and Applications of Semantics, Rhodes, Greece, Oct. 2019, Accessed: May 26, 2020. [Online]. Available: http://doras.dcu.ie/23801/.

[44] H. Pandit, R. Gachpaz Hamed, S. Lawless, and D. Lewis, "The Use of Open Data to Improve the Repeatability of Adaptivity and Personalisation Experiment," *24th Conference on User Modeling, Adaptation and Personalization*, p. 3, 2016.

[45] A. Hosseinzadeh Vahid, R. Gachpaz Hamed, and K. Koidl, "A Review of User-centred Information Retrieval Tasks," *24th Conference on User Modeling, Adaptation and Personalization*, p. 2, 2016.

[46] R. J. Torraco, "Writing integrative literature reviews: Guidelines and examples," *Human resource development review*, vol. 4, no. 3, pp. 356–367, 2005.

[47] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Information and software technology*, vol. 55, no. 12, pp. 2049–2075, 2013.

[48] R. Whittemore and K. Knafl, "The integrative review: updated methodology," *Journal of advanced nursing*, vol. 52, no. 5, pp. 546–553, 2005.

[49] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research*, vol. 104, pp. 333–339, 2019.

[50] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS quarterly*, pp. xiii–xxiii, 2002.

[51] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[52] Y. Wang and A. Kobsa, “Privacy-Enhancing Technologies,” *In Handbook of research on social and organizational liabilities in information security. (Vol. 5704, pp. 203–227)*, p. 23, 2009.

[53] H. Wang, L. Sun, and E. Bertino, “Building access control policy model for privacy preserving and testing policy conflicting problems,” *Journal of Computer and System Sciences*, vol. 80, no. 8, pp. 1493–1503, 2014.

[54] S. Kirrane, A. Mileo, and S. Decker, “Access control and the resource description framework: A survey,” *Semantic Web*, vol. 8, no. 2, pp. 311–352, 2017.

[55] P. Samarati and S. C. de Vimercati, “Access control: Policies, models, and mechanisms,” in *International School on Foundations of Security Analysis and Design*, 2000, pp. 137–196.

[56] E. Bertino and R. Sandhu, “Database security-concepts, approaches, and challenges,” *IEEE Transactions on Dependable and secure computing*, vol. 2, no. 1, pp. 2–19, 2005.

[57] R. S. Sandhu and P. Samarati, “Access control: principle and practice,” *IEEE communications magazine*, vol. 32, no. 9, pp. 40–48, 1994.

[58] M. Ennahbaoui and S. Elhajji, “Study of access control models,” in *Proceedings of the World Congress on Engineering*, 2013, vol. 2, pp. 3–5.

[59] J. Crampton, “On permissions, inheritance and role hierarchies,” in *Proceedings of the 10th ACM conference on Computer and communications security*, 2003, pp. 85–92.

[60] R. Sandhu, D. Ferraiolo, and R. Kuhn, “The NIST model for role-based access control: towards a unified standard,” in *ACM workshop on Role-based access control*, 2000, vol. 10, no. 344287.344301.

[61] A. Belokosztolszki, “Role-based access control policy administration,” University of Cambridge, Computer Laboratory, 2004.

[62] C. N. Zhang and C. Yang, “Designing a complete model of role-based access control system for distributed networks,” *J. Inf. Sci. Eng.*, vol. 18, no. 6, pp. 871–889, 2002.

[63] R. Ausanka-Crues, “Methods for access control: advances and limitations,” *Harvey Mudd College*, vol. 301, p. 20, 2001.

[64] D. R. Kuhn, E. J. Coyne, and T. R. Weil, “Adding attributes to role-based access control,” *Computer*, vol. 43, no. 6, pp. 79–81, 2010.

[65] W. Di, L. Jian, D. Yabo, and Z. Miaoliang, “Using semantic web technologies to specify constraints of RBAC,” in *Sixth International Conference on Parallel and Distributed Computing Applications and Technologies (PDCAT’05)*, 2005, pp. 543–545.

[66] T. Finin *et al.*, “R OWL BAC: representing role based access control in OWL,” in *Proceedings of the 13th ACM symposium on Access control models and technologies*, 2008, pp. 73–82.

[67] J. A. Calero, G. M. Perez, and A. G. Skarmeta, “Towards an authorisation model for distributed systems based on the Semantic Web,” *IET information security*, vol. 4, no. 4, pp. 411–421, 2010.

[68] V. C. Hu *et al.*, “Guide to attribute based access control (abac) definition and considerations (draft),” *NIST special publication*, vol. 800, no. 162, 2013.

[69] L. Cirio, I. F. Cruz, and R. Tamassia, “A role and attribute based access control system using semantic web technologies,” in *OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”*, 2007, pp. 1256–1266.

[70] I. F. Cruz, R. Gjomemo, B. Lin, and M. Orsini, “A location aware role and attribute based access control system,” in *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, 2008, pp. 1–2.

[71] I. F. Cruz, R. Gjomemo, B. Lin, and M. Orsini, “A constraint and attribute based security framework for dynamic role assignment in collaborative environments,” in *International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 2008, pp. 322–339.

[72] H. Shen, “A semantic-aware attribute-based access control model for web services,” in *International Conference on Algorithms and Architectures for Parallel Processing*, 2009, pp. 693–703.

[73] Z. He, L. Wu, H. Li, H. Lai, and Z. Hong, “Semantics-based Access Control Approach for Web Service.,” *JCP*, vol. 6, no. 6, pp. 1152–1161, 2011.

[74] T. Priebe, W. Dobmeier, and N. Kamprath, “Supporting attribute-based access control with ontologies,” in *First International conference on availability, reliability and security (ARES’06)*, 2006, pp. 8-pp.

[75] C. A. Ardagna *et al.*, “Enabling privacy-preserving credential-based access control with XACML and SAML,” in *2010 10th IEEE International Conference on Computer and Information Technology*, 2010, pp. 1090–1095.

[76] Z. Xu and S. D. Stoller, “Mining attribute-based access control policies,” *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 5, pp. 533–545, 2014.

[77] D. Lin, P. Rao, E. Bertino, N. Li, and J. Lobo, “EXAM: a comprehensive environment for the analysis of access control policies,” *International Journal of Information Security*, vol. 9, no. 4, pp. 253–273, 2010.

[78] J. Li and B. Zhang, “An ontology-based approach to improve access policy administration of attribute-based access control,” *International Journal of Information and Computer Security*, vol. 11, no. 4–5, pp. 391–412, 2019.

[79] M. J. Covington, W. Long, S. Srinivasan, A. K. Dev, M. Ahamad, and G. D. Abowd, “Securing context-aware applications using environment roles,” in *Proceedings of the sixth ACM symposium on Access control models and technologies*, 2001, pp. 10–20.

[80] G. Neumann and M. Strembeck, “An approach to engineer and enforce context constraints in an RBAC environment,” in *Proceedings of the eighth ACM symposium on Access control models and technologies*, 2003, pp. 65–79.

[81] P. McDaniel, “On context in authorization policy,” in *Proceedings of the eighth ACM symposium on Access control models and technologies*, 2003, pp. 80–89.

[82] G. K. Mostéfaoui and P. Brézillon, “A generic framework for context-based distributed authorizations,” in *International and Interdisciplinary Conference on Modeling and Using Context*, 2003, pp. 204–217.

[83] A. Corradi, R. Montanari, and D. Tibaldi, “Context-based access control for ubiquitous service provisioning,” in *Proceedings of the 28th Annual International Computer Software and Applications Conference, 2004. COMPSAC 2004.*, 2004, pp. 444–451.

[84] A. Toninelli, R. Montanari, L. Kagal, and O. Lassila, “A semantic context-aware access control framework for secure collaborations in pervasive computing environments,” in *International semantic web conference*, 2006, pp. 473–486.

[85] H. Shen and Y. Cheng, “A semantic context-based model for mobile web services access control,” *International Journal of Computer Network and Information Security*, vol. 3, no. 1, p. 18, 2011.

[86] A. S. M. Kayes, J. Han, and A. Colman, “An ontology-based approach to context-aware access control for software services,” in *International Conference on Web Information Systems Engineering*, 2013, pp. 410–420.

[87] A. Uszok *et al.*, *Policy and Contract Management for Semantic Web Services. to appear AAAI Spring Symposium*. Stanford University, California, USA, 2004.

[88] A. Uszok *et al.*, “KAoS policy management for semantic web services,” *IEEE Intelligent Systems*, vol. 19, no. 4, pp. 32–41, 2004.

[89] P. A. Bonatti and D. Olmedilla, “Rule-based policy representation and reasoning for the semantic web,” in *Reasoning Web International Summer School*, 2007, pp. 240–268.

[90] J. M. Bradshaw, S. Dutfield, P. Benoit, and J. D. Woolley, “KAoS: Toward an industrial-strength open agent architecture,” *Software agents*, vol. 13, pp. 375–418, 1997.

[91] J. Bradshaw *et al.*, “Representation and Reasoning for DAML-Based Policy and Domain Services in KAoS and Nomads,” *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, p. 9, 2003.

[92] A. Uszok, J. M. Bradshaw, R. Jeffers, A. Tate, and J. Dalton, “Applying KAoS Services to Ensure Policy Compliance for Semantic Web Services Workflow Composition and Enactment,” in *The Semantic Web – ISWC 2004*, vol. 3298, S. A. McIlraith, D. Plexousakis, and F. van Harmelen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 425–440.

[93] A. Uszok *et al.*, “DAML reality check: A case study of KAoS domain and policy services,” 2003.

[94] M. Schmidt-Schauß, “Subsumption in KL-ONE is Undecidable,” p. 17, 1989.

[95] A. Uszok *et al.*, “KAoS policy and domain services: toward a description-logic approach to policy representation, deconfliction, and enforcement,” in *Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks*, Lake Como, Italy, 2003, pp. 93–96, doi: 10.1109/POLICY.2003.1206963.

[96] L. Kagal, T. Finin, and A. Joshi, “A Policy Language for a Pervasive Computing Environment£,” *Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks. IEEE*, p. 12, 2003.

[97] L. Kagal, T. Finin, and A. Joshi, “A Policy Based Approach to Security for the Semantic Web,” Jan. 2003, pp. 402–418, doi: 10.1007/978-3-540-39718-2_26.

[98] L. Kagal, “A Policy-Based Approach to Governing Autonomous Behavior in Distributed Environments,” Sep. 2004, Accessed: Jan. 05, 2021. [Online]. Available: https://ebiquity.umbc.edu/paper/html/id/210/A-Policy-Based-Approach-to-Governing-Autonomous-Behavior-in-Distributed-Environments.

[99] G. Denker, L. Kagal, and T. Finin, “Security in the Semantic Web using OWL,” *Information Security Technical Report*, vol. 10, no. 1, pp. 51–58, Jan. 2005, doi: 10.1016/j.istr.2004.11.002.

[100] L. Kagal, “Using Semantic Web Technologies for Policy Management on the Web,” *Kagal, Lalana, et al. “Using semantic web technologies for policy management on the web.” Proceedings of the national conference on Artificial Intelligence. Vol. 21. No. 2. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press;*, p. 8, 2006.

[101] A. Toninelli, J. M. Bradshaw, L. Kagal, and R. Montanari, “Rule-based and Ontology-based Policies: Toward a Hybrid Approach to Control Agents in Pervasive Environments,” *Toninelli, Alessandra, et al. “Rule-based and*

ontology-based policies: Toward a hybrid approach to control agents in pervasive environments." proceedings of the Semantic Web and Policy Workshop, p. 14, 2005.

[102] A. Toninelli, R. Montanari, L. Kagal, and O. Lassila, "Proteus: A semantic context-aware adaptive policy model," in *Eighth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'07)*, 2007, pp. 129–140.

[103] R. Liscano and K. Wang, "A SIP-based architecture model for contextual coalition access control for ubiquitous computing," in *The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services*, San Diego, CA, USA, 2005, pp. 384–392, doi: 10.1109/MOBIQUITOUS.2005.8.

[104] A. Toninelli, A. Corradi, and R. Montanari, "A quality of context-aware approach to access control in pervasive environments," in *International Conference on Mobile Wireless Middleware, Operating Systems, and Applications*, 2009, pp. 236–251.

[105] A. Kitkowska, E. Wästlund, J. Meyer, and L. A. Martucci, "Is It Harmful? Re-examining Privacy Concerns," in *Privacy and Identity Management. The Smart Revolution: 12th IFIP WG 9.2, 9.5, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Ispra, Italy, September 4-8, 2017, Revised Selected Papers*, M. Hansen, E. Kosta, I. Nai-Fovino, and S. Fischer-Hübner, Eds. Cham: Springer International Publishing, 2018, pp. 59–75.

[106] N. Malhotra, S. Kim, and J. Agarwal, "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model," *Information Systems Research*, vol. 15, pp. 336–355, Dec. 2004, doi: 10.1287/isre.1040.0032.

[107] S. Kowalewski, M. Ziefle, J. Ziegeldorf, and K. Wehrle, "Like us on Facebook! – Analyzing User Preferences Regarding Privacy Settings in Germany," *Procedia Manufacturing*, vol. 3, pp. 815–822, Dec. 2015, doi: 10.1016/j.promfg.2015.07.336.

[108] Y. Shulman and J. Meyer, "Is Privacy Controllable?," in *Privacy and Identity Management. Fairness, Accountability, and Transparency in the Age of Big Data: 13th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Vienna, Austria, August 20-24, 2018, Revised Selected Papers*, E. Kosta, J. Pierson, D. Slamanig, S. Fischer-Hübner, and S. Krenn, Eds. Cham: Springer International Publishing, 2019, pp. 222–238.

[109] L. Kagal and H. Abelson, "Access control is an inadequate framework for privacy protection," *Proceedings of the W3C Privacy Workshop*, Jan. 2010.

[110] M. Hildebrandt, "D 7.12: Behavioural Biometric Profiling and," 2009.

[111] M. Hansen, "Marrying transparency tools with user-controlled identity management," in *IFIP International Summer School on the Future of Identity in the Information Society*, 2007, pp. 199–220.

[112] H. Hedbom, "A survey on transparency tools for enhancing privacy," in *IFIP Summer School on the Future of Identity in the Information Society*, 2008, pp. 67–82.

[113] M. Janic, J. P. Wijbenga, and T. Veugen, "Transparency Enhancing Tools (TETs): An Overview," in *2013 Third Workshop on Socio-Technical Aspects in Security and Trust*, Jun. 2013, pp. 18–25, doi: 10.1109/STAST.2013.11.

[114] M. Hildebrandt, "Profiling and AmI," in *The future of identity in the information society*, Springer, 2009, pp. 273–310.

[115] C. Zimmermann, “A Categorization of Transparency-Enhancing Technologies,” *arXiv:1507.04914 [cs]*, Jul. 2015.

[116] C. Zimmermann, R. Accorsi, and G. Müller, “Privacy dashboards: reconciling data-driven business models and privacy,” in *2014 Ninth International Conference on Availability, Reliability and Security*, 2014, pp. 152–157.

[117] R. Accorsi, C. Zimmermann, and G. Müller, “On taming the inference threat in social networks,” 2012.

[118] B. Schneier, “A taxonomy of social networking data,” *IEEE Security & Privacy*, vol. 8, no. 4, pp. 88–88, 2010.

[119] P. G. Kelley, P. Hankes Drielsma, N. Sadeh, and L. F. Cranor, “User-controllable learning of security and privacy policies,” in *Proceedings of the 1st ACM workshop on Workshop on AISec*, 2008, pp. 11–18.

[120] N. Sadeh *et al.*, “Understanding and capturing people’s privacy policies in a mobile social networking application,” *Personal and Ubiquitous Computing*, vol. 13, no. 6, pp. 401–412, 2009.

[121] J. Y. Tsai, P. Kelley, P. Drielsma, L. F. Cranor, J. Hong, and N. Sadeh, “Who’s viewed you? The impact of feedback in a mobile location-sharing application,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 2003–2012.

[122] E. Toch *et al.*, “Empirical models of privacy in location sharing,” in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 2010, pp. 129–138.

[123] R. Balebako, J. Jung, W. Lu, L. F. Cranor, and C. Nguyen, “‘ Little brothers watching you’ raising awareness of data leaks on smartphones,” in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, 2013, pp. 1–11.

[124] M. Y. Mun, D. H. Kim, K. Shilton, D. Estrin, M. Hansen, and R. Govindan, “PDVLoc: A personal data vault for controlled location data sharing,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 10, no. 4, pp. 1–29, 2014.

[125] R. Schlegel, A. Kapadia, and A. J. Lee, “Eyeing your exposure: quantifying and controlling information sharing for improved privacy,” in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, 2011, pp. 1–14.

[126] I. Bilogrevic, K. Huguenin, B. Agir, M. Jadliwala, and J.-P. Hubaux, “Adaptive information-sharing for privacy-aware mobile social networks,” in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, 2013, pp. 657–666.

[127] D. Biswas, I. Aad, and G. P. Perrucci, “Privacy panel: Usable and quantifiable mobile privacy,” in *2013 International Conference on Availability, Reliability and Security*, 2013, pp. 218–223.

[128] K. Abdullah, G. Conti, and R. Beyah, “A visualization framework for self-monitoring of web-based information disclosure,” in *2008 IEEE International Conference on Communications*, 2008, pp. 1700–1707.

[129] C. Louw and S. von Solms, “Personally identifiable information leakage through online social networks,” in *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, 2013, pp. 68–71.

[130] S. Trabelsi and J. Sendor, “Sticky policies for data control in the cloud,” in *2012 Tenth Annual International Conference on Privacy, Security and Trust*, 2012, pp. 75–80.

[131] J. Kolter, M. Netter, and G. Pernul, "Visualizing past personal data disclosures," in *2010 International Conference on Availability, Reliability and Security*, 2010, pp. 131–139.

[132] C. Riederer, D. Echickson, S. Huang, and A. Chaintreau, "Findyou: A personal location privacy auditing tool," in *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, pp. 243–246.

[133] T. pulls, *The Data Track , Available: https://github.com/pylls/datatrack*. 2016.

[134] G. Hsieh, K. P. Tang, W. Y. Low, and J. I. Hong, "Field deployment of IMBuddy: a study of privacy control and feedback mechanisms for contextual IM," in *International Conference on Ubiquitous Computing*, 2007, pp. 91–108.

[135] Oxford Dictionaries, "Effectiveness | Definition of Effectiveness by Oxford Dictionary on Lexico.com also meaning of Effectiveness," *Lexico Dictionaries | English*. https://www.lexico.com/definition/effectiveness (accessed Jan. 11, 2021).

[136] Int. Org. Standardization, Geneva, Switzerland, Tech, "ISO 9241-210:2010 Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems," *ISO*. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/20/52075.html.

[137] C. Bier, K. Kühne, and J. Beyerer, "PrivacyInsight: The Next Generation Privacy Dashboard," in *Privacy Technologies and Policy*, Cham, 2016, pp. 135–152, doi: 10.1007/978-3-319-44760-5_9.

[138] J. Angulo, S. Fischer-Hübner, T. Pulls, and E. Wästlund, "Usable Transparency with the Data Track: A Tool for Visualizing Data Disclosures," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, Apr. 2015, pp. 1803–1808, doi: 10.1145/2702613.2732701.

[139] Standardization, Geneva, Switzerland, Tech., "ISO 9241-110:2006 Ergonomics of human-system interaction — Part 110: Dialogue principles," *ISO*. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/03/80/38009.html.

[140] A. S. Patrick and S. Kenny, "From Privacy Legislation to Interface Design: Implementing Information Privacy in Human-Computer Interactions," in *Privacy Enhancing Technologies*, Berlin, Heidelberg, 2003, pp. 107–124, doi: 10.1007/978-3-540-40956-4_8.

[141] C. Hao, *A Survey and Classification of Privacy-Preservation Mechanisms for Cloud Data Management*. 2014.

[142] J. Domingo-Ferrer, O. Farràs, J. Ribes-González, and D. Sánchez, "Privacy-preserving cloud computing on sensitive data: A survey of methods, products and challenges," *Computer Communications*, vol. 140–141, pp. 38–60, May 2019, doi: 10.1016/j.comcom.2019.04.011.

[143] L. Vegh, "A Survey of Privacy and Security Issues for the Internet of Things in the GDPR Era," in *2018 International Conference on Communications (COMM)*, Jun. 2018, pp. 453–458, doi: 10.1109/ICComm.2018.8484769.

[144] M. Seliem, K. Elgazzar, and K. Khalil, "Towards Privacy Preserving IoT Environments: A Survey," 2018. https://www.hindawi.com/journals/wcmc/2018/1032761/(accessed Jan. 11, 2021).

[145] S. Wachter, “The GDPR and the Internet of Things: a three-step transparency model,” *Law, Innovation and Technology*, vol. 10, no. 2, pp. 266–294, Jul. 2018, doi: 10.1080/17579961.2018.1527479.

[146] E. Kani-Zabihi and M. Helmhout, “Increasing service users’ privacy awareness by introducing on-line interactive privacy features,” in *Information Security Technology for Applications. NordSec 2011.*, vol. 7161, Heidelberg, Germany: Springer, 2012, pp. 131–148.

[147] A. Popescu *et al.*, “Increasing Transparency and Privacy for Online Social Network Users – USEMP Value Model, Scoring Framework and Legal,” in *Privacy Technologies and Policy*, Cham, 2016, pp. 38–59, doi: 10.1007/978-3-319-31456-3_3.

[148] A. Zavou, V. Pappas, V. P. Kemerlis, M. Polychronakis, G. Portokalidis, and A. D. Keromytis, “Cloudopsy: An Autopsy of Data Flows in the Cloud,” in *Human Aspects of Information Security, Privacy, and Trust*, Berlin, Heidelberg, 2013, pp. 366–375, doi: 10.1007/978-3-642-39345-7_39.

[149] M. Pistoia, O. Tripp, P. Centonze, and J. W. Ligman, “Labyrinth: Visually Configurable Data-Leakage Detection in Mobile Applications,” in *2015 16th IEEE International Conference on Mobile Data Management*, Jun. 2015, vol. 1, pp. 279–286, doi: 10.1109/MDM.2015.69.

[150] P. Kruchten, “The 4+1 View Model of Architecture,” *IEEE Software*, vol. 12, pp. 45–50, Nov. 1995, doi: 10.1109/52.469759.

[151] A. Adadi and M. Berrada, “Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI),” *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[152] T. Lombrozo and N. Z. Gwynne, “Explanation and inference: Mechanistic and functional explanations guide property generalization,” *Frontiers in Human Neuroscience*, vol. 8, p. 700, 2014.

[153] S. Krening, B. Harrison, K. M. Feigh, C. L. Isbell, M. Riedl, and A. Thomaz, “Learning from explanations using sentiment and advice in RL,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 1, pp. 44–55, 2016.

[154] A. Páez, “The pragmatic turn in explainable artificial intelligence (XAI),” *Minds and Machines*, vol. 29, no. 3, pp. 441–459, 2019.

[155] F. Michel, L. Djimenou, C. F. Zucker, and J. Montagnat, “xR2RML: Relational and non-relational databases to RDF mapping language,” 2017.

[156] M. Hert, G. Reif, and H. C. Gall, “A comparison of RDB-to-RDF mapping languages,” in *Proceedings of the 7th International Conference on Semantic Systems*, 2011, pp. 25–32.

[157] O. Drozd and S. Kirrane, “Privacy CURE: Consent Comprehension Made Easy,” 2020.

[158] N. Shen *et al.*, “Understanding the patient privacy perspective on health information exchange: a systematic review,” *International journal of medical informatics*, vol. 125, pp. 1–12, 2019.

[159] D. Pierina Brustolin Spagnuelo, “Defining, Measuring, and Enabling Transparency for Electronic Medical Systems,” PhD Thesis, University of Luxembourg, Luxembourg, 2018.

[160] S. Univesity, *Protege: a free, open source ontology editor and knowledge-base framework.* Stanford Center for Biomedical Informatics Research, Stanford, CA, USA.

[161] M. Bazire and P. Brézillon, “Understanding context before using it,” in *International and Interdisciplinary Conference on Modeling and Using Context*, 2005, pp. 29–40.

[162] A. K. Dey, “Understanding and using context,” *Personal and ubiquitous computing*, vol. 5, no. 1, pp. 4–7, 2001.

[163] M. Horridge and S. Bechhofer, “The OWL API: a Java API for working with OWL 2 ontologies,” in *Proceedings of the 6th International Conference on OWL: Experiences and Directions-Volume 529*, 2009, pp. 49–58.

[164] M. Horridge, B. Parsia, and U. Sattler, “Justification oriented proofs in OWL,” in *International Semantic Web Conference*, 2010, pp. 354–369.

[165] M. Horridge, N. Drummond, J. Goodwin, A. L. Rector, R. Stevens, and H. Wang, “The Manchester OWL syntax.,” in *OWLed*, 2006, vol. 216.

[166] M. Horridge, B. Parsia, and U. Sattler, “Laconic and precise justifications in OWL,” in *International semantic web conference*, 2008, pp. 323–338.

[167] N. E. Fuchs and K. Kaljurand, “Attempto Controlled English meets the challenges of knowledge representation, reasoning, interoperability and user interfaces,” 2006.

[168] K. Kaljurand and N. E. Fuchs, “Verbalizing owl in attempto controlled english,” 2007.

[169] E. Gansner, E. Koutsofios, and S. North, *Drawing graphs with dot*. Technical report, AT&T Research. URL http://www. graphviz. org/Documentation …, 2006.

[170] M. Horridge, *Justification based explanation in ontologies*. The University of Manchester (United Kingdom), 2011.

[171] D. de Paula, D. Saraiva, R. Natália, N. Garcia, and V. Leithardt, “Study of a Context Quality Model for UbiPri Middleware,” *KnE Engineering*, pp. 523–538, 2020.

[172] O. Biran and C. Cotton, “Explanation and justification in machine learning: A survey,” in *IJCAI-17 workshop on explainable AI (XAI)*, 2017, vol. 8, no. 1, pp. 8–13.

[173] F. K. Došilović, M. Brčić, and N. Hlupić, “Explainable artificial intelligence: A survey,” in *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*, 2018, pp. 0210–0215.

[174] D. Gunning and D. W. Aha, “DARPA’s explainable artificial intelligence program,” *AI Magazine*, vol. 40, no. 2, pp. 44–58, 2019.

[175] T. Miller, P. Howe, and L. Sonenberg, “Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences,” *arXiv preprint arXiv:1712.00547*, 2017.

[176] I. Nunes and D. Jannach, “A systematic review and taxonomy of explanations in decision support and recommender systems,” *User Modeling and User-Adapted Interaction*, vol. 27, no. 3–5, pp. 393–444, 2017.

[177] F. Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Vaughan, and H. Wallach, “Manipulating and measuring model interpretability,” *arXiv preprint arXiv:1802.07810*, 2018.

[178] R. M. Byrne, “The construction of explanations,” in *AI and Cognitive Science’90*, Springer, 1991, pp. 337–351.

[179] P. Thagard, “Explanatory coherence,” *Behavioral and brain sciences*, vol. 12, no. 3, pp. 435–502, 1989.

[180] J. Y. Halpern and J. Pearl, “Causes and explanations: A structural-model approach. Part II: Explanations,” *The British journal for the philosophy of science*, vol. 56, no. 4, pp. 889–911, 2005.

[181] J. Pearl, “Causal inference,” *Causality: Objectives and Assessment*, pp. 39–58, 2010.

[182] D. J. Hilton, “Conversational processes and causal explanation.,” *Psychological Bulletin*, vol. 107, no. 1, p. 65, 1990.

[183] D. J. Hilton, “Mental models and causal explanation: Judgements of probable cause and explanatory relevance,” *Thinking & Reasoning*, vol. 2, no. 4, pp. 273–308, 1996.

[184] J. M. Jaspars and D. J. Hilton, “Mental models of causal reasoning.,” 1988.

[185] H. P. Grice, “Logic and conversation,” in *Speech acts*, Brill, 1975, pp. 41–58.

[186] J. McClure, “Goal-based explanations of actions and outcomes,” *European review of social psychology*, vol. 12, no. 1, pp. 201–235, 2002.

[187] S. J. Read and A. Marcus-Newhall, “Explanatory coherence in social explanations: A parallel distributed processing account.,” *Journal of Personality and Social Psychology*, vol. 65, no. 3, p. 429, 1993.

[188] A. Tversky and D. Kahneman, “Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment.,” *Psychological review*, vol. 90, no. 4, p. 293, 1983.

[189] M. Chromik and M. Schuessler, “A Taxonomy for Human Subject Evaluation of Black-Box Explanations in XAI.,” 2020.

[190] F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” *arXiv preprint arXiv:1702.08608*, 2017.

[191] P. Antunes, V. Herskovic, S. F. Ochoa, and J. A. Pino, “Structuring dimensions for collaborative systems evaluation,” *ACM computing surveys (CSUR)*, vol. 44, no. 2, pp. 1–28, 2008.

[192] H. Lakkaraju, S. H. Bach, and J. Leskovec, “Interpretable decision sets: A joint framework for description and prediction,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1675–1684.

[193] S. Mohseni and E. D. Ragan, “A human-grounded evaluation benchmark for local explanations of machine learning,” *arXiv preprint arXiv:1801.05075*, 2018.

[194] F. Yang, M. Du, and X. Hu, “Evaluating explanation without ground truth in interpretable machine learning,” *arXiv preprint arXiv:1907.06831*, 2019.

[195] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, “Metrics for explainable AI: Challenges and prospects,” *arXiv preprint arXiv:1812.04608*, 2018.

[196] P. Murmann, “Towards Usable Transparency via Individualisation,” *Diss. Karlstads universitet*, 2019.

[197] A. Holzinger, A. Carrington, and H. Müller, “Measuring the quality of explanations: the system causability scale (SCS),” *KI-Künstliche Intelligenz*, pp. 1–6, 2020.

[198] A. M. Surprenant and I. Neath, *Principles of memory*. Psychology Press, 2013.

[199] S. Mohseni, N. Zarei, and E. D. Ragan, “A survey of evaluation methods and measures for interpretable machine learning,” *arXiv preprint arXiv:1811.11839*, 2018.

[200] M. Narayanan, E. Chen, J. He, B. Kim, S. Gershman, and F. Doshi-Velez, “How do humans understand explanations from machine learning systems? an evaluation of the human-interpretability of explanation,” *arXiv preprint arXiv:1802.00682*, 2018.

[201] R. Porst, P. Schmidt, and K. Zeifang, “Comparisons of subgroups by models with multiple indicators,” *Sociological Methods & Research*, vol. 15, no. 3, pp. 303–315, 1987.

[202] E. Singer, “The use of incentives to reduce nonresponse in household surveys,” *Survey nonresponse*, vol. 51, pp. 163–177, 2002.

[203] D. H. Sova and J. Nielsen, *234 tips and tricks for recruiting users as participants in usability studies*. Nielsen Norman Group, 2010.

[204] R. B. Johnson and L. Christensen, *Educational research: Quantitative, qualitative, and mixed approaches*. SAGE Publications, Incorporated, 2019.

[205] M. Coughlan, P. Cronin, and F. Ryan, “Step-by-step guide to critiquing research. Part 1: quantitative research,” *British journal of nursing*, vol. 16, no. 11, pp. 658–663, 2007.

[206] A. William and T. Tullis, *Measuring the User Experience:: collecting, analyzing, and presenting usability metrics*. Elsevier, 2013.

[207] L. Clayton, *Using the” thinking-aloud” method in cognitive interface design*. Yorktown Heights, NY: IBM TJ Watson Research Center, 1982.

[208] “ISO 9241-11:1998(en), Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability.” https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en.

[209] A. Barredo Arrieta *et al.*, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[210] J. Salminen, J. M. Santos, S.-G. Jung, M. Eslami, and B. J. Jansen, “Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas,” *International Journal of Human–Computer Interaction*, vol. 36, no. 8, pp. 788–800, May 2020, doi: 10.1080/10447318.2019.1688946.

[211] N. Grgić-Hlača, C. Engel, and K. P. Gummadi, “Human decision making with machine assistance: An experiment on bailing and jailing,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–25, 2019.

[212] A. Braca, B. Spillane, V. Wade, and P. Dondio, “Pilot Data Collection Survey and Analytical Techniques for Persuasion Engineering Systems.”

[213] B. Spillane, S. Lawless, and V. Wade, “Increasing and decreasing perceived bias by distorting the quality of news website design,” in *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*, 2018, pp. 1–13.

[214] J. Sauro and J. R. Lewis, *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016.

[215] A. Bangor, P. T. Kortum, and J. T. Miller, “An empirical evaluation of the system usability scale,” *Intl. Journal of Human–Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008.

[216] J. R. Lewis and J. Sauro, “The factor structure of the system usability scale,” in *International conference on human centered design*, 2009, pp. 94–103.

[217] A. Bangor, P. Kortum, and J. Miller, “Determining what individual SUS scores mean: Adding an adjective rating scale,” *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009.

[218] N. M. Lucey, “More than meets the I: User-satisfaction of computer systems,” *Unpublished thesis for Diploma in Applied Psychology, University College Cork, Cork, Ireland*, 1991.

[219] J. R. Lewis, “IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use,” *International Journal of Human-Computer Interaction*, vol. 7, no. 1, pp. 57–78, 1995.

[220] E. R. Van Teijlingen and V. Hundley, “The importance of pilot studies,” 2001.

[221] D. A. Dillman, *Mail and Internet surveys: The tailored design method–2007 Update with new Internet, visual, and mixed-mode guide*. John Wiley & Sons, 2011.

[222] C. M. Macdonald, *Understanding Usefulness in Human-Computer Interaction to Enhance User Experience Evaluation.* ERIC, 2012.

[223] J. Grudin, “Utility and usability: research issues and development contexts,” *Interacting with computers*, vol. 4, no. 2, pp. 209–217, 1992.

[224] W. J. Orlikowski and D. C. Gash, “Technological frames: making sense of information technology in organizations,” *ACM Transactions on Information Systems (TOIS)*, vol. 12, no. 2, pp. 174–207, 1994.

[225] V. Venkatesh, F. Davis, and M. G. Morris, “Dead or alive? The development, trajectory and future of technology adoption research.,” *Journal of the association for information systems*, vol. 8, no. 4, p. 1, 2007.

[226] D. Doyle, P. Cunningham, and P. Walsh, “An Evaluation of The Usefulness Of Explanation In A Case-Based Reasoning System For Decision Support In Bronchiolitis Treatment,” *Computational Intelligence*, vol. 22, no. 3–4, pp. 269–281, 2006.

[227] S. Coppers *et al.*, “Intellingo: An Intelligible Translation Environment,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–13.

[228] M. A. Albanese and D. L. Sabers, “Multiple true-false items: a study of interitem correlations, scoring alternatives, and reliability estimation,” *Journal of Educational Measurement*, vol. 25, no. 2, pp. 111–123, 1988.

[229] N. I. Siddiqui, V. H. Bhavsar, A. V. Bhavsar, and S. Bose, “Contemplation on marking scheme for Type X multiple choice questions, and an illustration of a practically applicable scheme,” *Indian journal of pharmacology*, vol. 48, no. 2, p. 114, 2016.

[230] J. Sauro and J. R. Lewis, “Correlations among prototypical usability metrics: evidence for the construct of usability,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 1609–1618.

[231] I. Lage *et al.*, “Human evaluation of models built for interpretability,” in *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 2019, vol. 7, no. 1, pp. 59–67.

[232] M. Bilgic and R. J. Mooney, “Explaining recommendations: Satisfaction vs. promotion,” in *Beyond Personalization Workshop, IUI*, 2005, vol. 5, p. 153.

[233] S. Mohseni, N. Zarei, and E. D. Ragan, “A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems,” *arXiv*, p. arXiv-1811, 2018.

[234] N. Tintarev and J. Masthoff, “Evaluating the effectiveness of explanations for recommender systems,” *User Modeling and User-Adapted Interaction*, vol. 22, no. 4–5, pp. 399–439, 2012.

[235] L. Rozenblit and F. Keil, “The misunderstood limits of folk science: An illusion of explanatory depth,” *Cognitive science*, vol. 26, no. 5, pp. 521–562, 2002.

[236] J. F. Yates, J.-W. Lee, and J. G. Bush, “General knowledge overconfidence: cross-national variations, response style, and ‘reality,’” *Organizational behavior and human decision processes*, vol. 70, no. 2, pp. 87–94, 1997.

[237] J. Kruger and D. Dunning, “Unskilled and unaware of it: how difficulties in recognizing one’s own incompetence lead to inflated self-assessments.,” *Journal of personality and social psychology*, vol. 77, no. 6, p. 1121, 1999.

[238] F. C. Keil, “Explanation and understanding,” *Annu. Rev. Psychol.*, vol. 57, pp. 227–254, 2006.

[239] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, “The NeOn methodology for ontology engineering,” in *Ontology engineering in a networked world*, Springer, 2012, pp. 9–34.

[240] P. A. Bonatti, S. Kirrane, I. M. Petrova, L. Sauro, and E. Schlehahn, “The SPECIAL usage policy language,” V0. 1. Tech. rep, 2018. [Online]. Available: https://www.specialprivacy.eu/vocabs.

[241] M. Lizar and D. Turner, “Consent Receipt Specification, version 1.0. 0,” *March. https://kantarainitiative. org/confluence/display/info sharing/Home*, 2017.

[242] B. A. Bonatti, W. Dullaert, J. D. Fernandez, S. Kirrane, U. Milosevic, and A. Polleres, *The SPECIAL Policy Log Vocabulary (Nov 2018)*.

.

Appendix A. Vocabulary for Data Privacy

Creating a Vocabulary for Data Privacy

During the design of the XPDA service and the implementation of its prototype in the motivating scenario, we encountered a gap related to lack of standard and agreed-upon vocabulary to:

- describe and interchange personal data being relevant to support the rights of data owner mentioned in chapter 3 (Articles 12-23) of the GDPR [25].
- describe the purposes of personal data handling and categories of processing to comply with the required legal bases such as consent
- align the terminology of privacy legislation - such as the GDPR, to allow organisations to claim compliance with such regulations using machine-readable information.

This chapter describes our contribution in a W3C Community Group (CG), named as ‘Data Privacy Vocabularies and Controls Community Group (DPVCG)’, to address these challenges. Also, Data Privacy Vocabulary (DPV), as a comprehensive, standardised set of terms of personal data handling and privacy policies annotation, would be described as an outcome of this contribution.

- **Data Privacy Vocabularies and Controls Community Group**

This W3C Community Group (CG) was formally established on 25th May 2018 - the implementation date of the GDPR. The group has more than 50 members to date representing academia, industry, legal experts, and other stakeholders. General information about the group along with the meetings’ notes and resources publicly available through a wiki page[13] and its discussions are open via the public mailing list[14]. After over a year of a collaborative effort, the ‘Data Privacy Vocabulary’ (DPV) was published on 25th July 2019.

- **Methodology**

[13] https://www.w3.org/community/dpvcg/wiki/Main_Page

[14] https://lists.w3.org/Archives/Public/public-dpvcg/

Existing relevant use cases and vocabularies were collected and documented in a wiki document[15] through individual submissions by CG members. Relevant terms were then identified from each vocabulary. Then, their relevance, requirements, and applicable use cases categorised as various taxonomies and listed in the wiki page. Along with working on the categorisation of the terms, the need for developing an ontology is realised in order to represent relations between these terms. The process of ontology development adopted the NeOn methodology scenarios [239] and SPECIAL Usage Policy Language [240] as basic instruments. Top-level concepts and their hierarchies were proposed, discussed in several co-creation sessions, and added in a collaborative spreadsheet hosted on the Google Sheets platform[16]. The vocabulary was created through a script[17] that extracted terms using the Google Drive API and generated RDF serialisations and documentation using rdflib[18] and ReSpec[19], respectively.

- **Data Privacy Vocabulary**

As a result of the process above, the 'Data Privacy Vocabulary' (DPV) has been published on 25th July 2019 at a namespace[20] as a public draft for feedback. The current vocabulary provides terms (classes and properties) to annotate and categorise instances of legally compliant personal data handling. In particular, DPV provides extensible concepts and relationships to describe the following components:

1. Personal Data Categories
2. Purposes
3. Processing Categories
4. Technical and Organisational Measures
5. Legal Basis
6. Consent
7. Recipients, Data Controllers, Data Subjects

These components are intended to express Personal Data Handling in a machine-readable form by specifying the personal data categories undergoing some processing,

[15] https://www.w3.org/community/dpvcg/wiki/Taxonomy

[16] https://www.google.com/sheets/about/

[17] https://github.com/dpvcg/extract-sheets/

[18] https://github.com/RDFLib/rdflib

[19] https://github.com/w3c/respec

[20] http://w3.org/ns/dpv

for some purpose, by the data controller, justified by legal basis, with specific technical and organisational measures, which may result in data being shared with some recipient. The vocabulary is built up in a modular fashion, where each 'module' covers one of the above-listed aspects, and which is linked together using a core Base Vocabulary.

Base Ontology and Personal Data Category module would be described in next sections because of active participation of the primary author of this thesis in their development and others were outlined and can be found for further reading in [43].

- **Base Ontology**

The 'Base Ontology' describes the top-level classes defining a policy for legal personal data handling. Sub-Classes and properties for each top-level class are further elaborated using sub-vocabularies, which are available as separate modules and are outlined in [43]. The modular approach to provide a separate base ontology makes it possible to use sub-vocabularies by sharing *dpv: namespace*. The core concepts of the Base Ontology module and their relationships are depicted in Figure A.1.

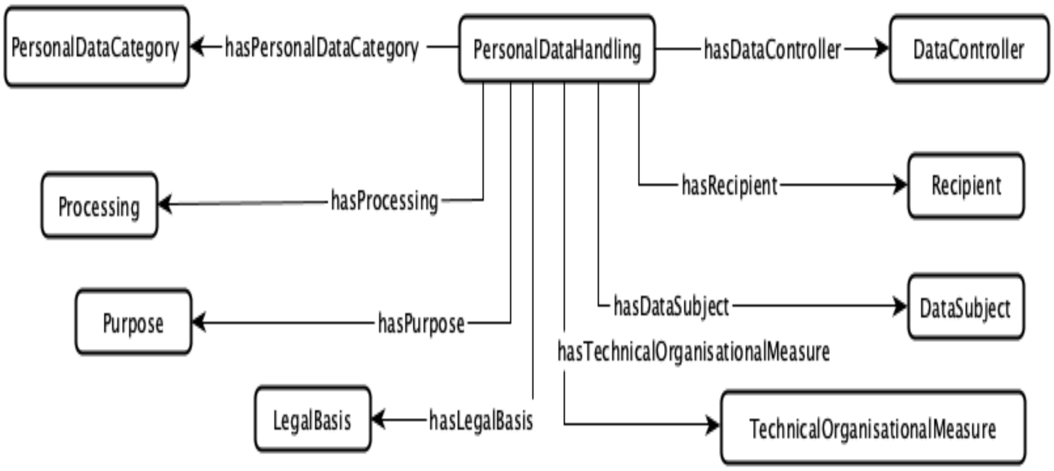


Figure 0A.1. DPV base Ontology classes and properties

- **Personal Data Categories**

DPV adopts its broad top-level personal data categories from the taxonomy provided by EnterPrivacy[21]. The top-level concepts in this taxonomy refer to the nature of information (financial, social, tracking) and its inherent source (internal, external). Each top-level concept is represented in the DPV as a class and is further elaborated by subclasses for referring to specific categories of information - such as preferences or demographics. Inspired from the motivating scenario of this thesis, the class dpv:SpecialCategoryOfPersonalData is defined to represent categories that are 'special' or 'sensitive' and require additional conditions according to Article 9 of GDPR.

The categories defined in the personal data taxonomy can be used directly or further extended to refer to the scope of personal data used in processing. The taxonomy can be extended by defining the subclasses of respective classes to depict specialised concepts or combined with classes to indicate specific contexts. The class dpv:DerivedPersonalData is an example of such context where information has been inferred from existing information of opinions from social media. While the taxonomy is by no means exhaustive, the aim is to provide sufficient coverage of abstract categories of personal data which can be extended using the subclass mechanism to represent concepts used in the real world. For instance, Figure A.2 shows the hierarchy of concepts for classifying depictions of individuals in pictures.

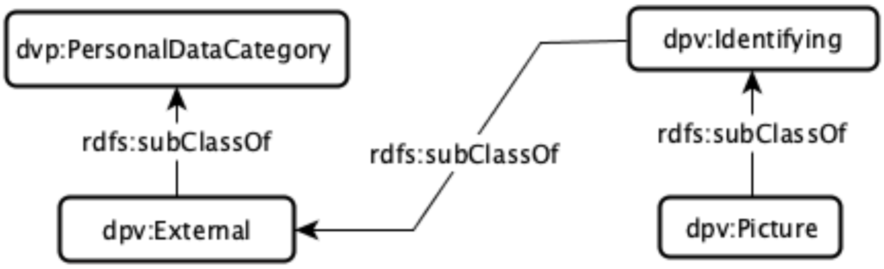


Figure 0.2. Hierarchy of concepts for classifying depictions of individuals in picture (inspired by EnterPrivacy)

[21] https://enterprivacy.com/2017/03/01/categories-of-personal-information/

- **Potential Adoption and Usage**

The primary aim of DPV is to assist in the representation of information concerning privacy in the context of personal data processing. To this end, it models concepts at an abstract or top-level to cover a broad range of concepts. This shall enable the DPV to be used as a domain-independent vocabulary which can be extended or specialised for specific domains or use cases. Though the DPV does not define or restrict how such an extension should be created, this section highlights some suggested methods for its adoption and usage. Firstly, the modular nature of DPV enables the adoption of a selected subset of the vocabulary only to address a specific use case. For example, an adopter may only wish to deploy the concepts under Purpose and PersonalDataCategory without using/describing all aspects of a particular PersonalDataHandling from the base vocabulary.

Besides, the use of RDFS and OWL enables extending the DPV in a compatible manner to define domain-specific use cases. For example, an extension targeting the finance domain can define additional concepts by using RDFS' subclass mechanism. Such an extension, when represented as an ontology, will be compatible with the DPV and will enable semantic interoperability of information, and ideally applications such as automated compliance checking for privacy policies and data handling records annotated with DPV and its extensions. Nevertheless, the DPV is intended to be used as an interoperable vocabulary where terms are structured in a hierarchy and have an unambiguous definition to enable common agreement over their semantics which can limit the flexibility of adopting these concepts to other pre-defined vocabularies, as seen in the case of Consent Receipts [241] and the SPECIAL vocabularies [242].

Appendix B. Informed consent

TRINITY COLLEGE DUBLIN

INFORMED CONSENT FORM

Thank you for your interest in taking part in this experiment. The information below is provided to inform your decision about consenting to take part in this experiment.

Researcher: Roghaiyeh(Ramisa) Gachpaz Hamed

Background of the research:

Personal data has been described as the new oil of the Internet and a new currency of the digital world. The digital age has created new ways of collecting, accessing, analysing and using these data, often across different jurisdictions. Such extensive collection provokes privacy concerns on the part of users. On the 28th birthday of World Wide Web in March 2017 "Losing control over personal data" was expressed as one of the three biggest challenges facing the web.

Despite the publication of legislation in multiple countries to preserve privacy, there is still a need to better inform users when sharing their personal data. Users need to be provided with an effective way to help them during the process of disclosing their personal data. They need to be assisted so that they have more control over their data when services and companies are accessing it.

This research proposes an approach for empowering users to understand access to their personal data. It focusses on finding easy and effective approaches (textual and visual) for explaining complex decisions to the user during disclosure of their data.

Aim:

The aim of this questionnaire is to gather information about comprehension of output from a system. The output is an explanation that shows the reason for a disclosure decision over personal data.

Procedure:

If you agree to participate in the research, you will be asked to complete a questionnaire in which you will answer a series of multiple-choice questions regarding the topics outlined above.

The experiment should take about 20-25 minutes where:

- Basic information about the experiment will be presented.
- You will confirm your consent to being involved in this experiment.
- You will be asked to fill out some background information.

This will include questions on:

- Physical characteristics, age range and sex.
- Educational background
- English reading skill
- Importance of being informed about access to personal data
- Background knowledge about graphs

- You will be asked to work as below for **two** test cases:
  - To read a very short scenario and output of the system.
  - To answer questions about the output of the system.

On completion of the experiment you will be returned to the Prolific Academic website

Publication:

The data gathered from this questionnaire will be used as part of the researcher's PhD thesis and may be also be presented at academic conferences. All information you provide will be treated with full confidentiality and, if published, will not be identifiable as yours.

Declaration:

- I am at least 18 years of age and competent to provide consent.
- I agree to participate in this experiment for the stated monitory reward advertised on the Prolific Academic Crowdsourcing marketplace.
- I have read a document providing information about this research and this consent form. I understand to my satisfaction, the description of the research that has been provided to me.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.

- I understand that none of the questions are mandatory except for the Prolific Academic ID question which is necessary to receive payment.
- I understand that should I complete the experiment, I may subsequently ask that my submission be removed from the study and any records destroyed.
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that my participation is fully anonymous and that only generic details about me are being recorded, none of which can be used to identify me at a later stage.
- I will not name third parties in any open text field of the questionnaire.
- I understand that if I, or anyone in my family has a history of epilepsy, then I am proceeding at my own risk.
- I understand that my completed submission will be reviewed to ascertain that a serious attempt has been made to answer the questions and fulfil the requirements of the experiment. Submissions which fail certain attention questions or provide obviously erroneous information will be discarded and payment will not be forwarded.
- I understand that the experiment is hosted by Trinity College Dublin and that the researcher will hold my anonymous submission in the same institution.
- All participants are eligible for the reward as advertised on the Prolific Academic crowdsourcing marketplace website. I understand I have the right to withdraw from the experiment at any stage, but I will not be eligible to claim the reward.
- I understand that attention questions will be used in the experiment and failure to answer one or more of these correctly will result in my submission being rejected and payment not being delivered.
- I understand that this is a crowdsourced experiment and that there is no supervision or guidance. I understand that I will not be able to ask questions and

that if I have any doubts or hesitation about participation that I should not continue.

Print:

To print a copy of this Informed Consent Form please use File > Print from the corner of your browser. Alternatively, you may email the researcher below to request a digital copy.

Research group:

This study is being undertaken by researchers in the ADAPT Centre in Trinity College Dublin. The SFI funded ADAPT project is a dynamic Academia-Industry partnership with over 100 researchers developing novel technologies addressing the key localisation challenges of volume, access and personalisation. If you would like further details about the study, feel free to contact Roghaiyeh(Ramisa) Gachpaz Hamed in Trinity College Dublin.

This research is supported by the Science Foundation Ireland and the ADAPT Centre at Trinity College, Dublin.

Statement of researcher's responsibility:

I, the lead researcher, have provided the above information and the information contained on the Participant Information Sheet in good faith and I commit to abide by it.

Roghaiyeh(Ramisa) Gachpaz Hamed
ADAPT Centre
Department of Computer Science
Trinity College Dublin
Email: ramisa.hamed@adaptcentre.ie
Phone: + 353 1 896 1765

Appendix C. Supplementary Materials

The data and code for implementing the XPDA service prototype and the results of the user study, discussed in Chapter 5, are available in a private GitHub repository[22]. Further, the online questionnaires of the user study are available through Qualtrics[23]. Please contact the author of this thesis on Ramisa.Hamed@tcd.ie to access either the repository or the questionnaires.

[22] https://github.com/RamisaHamed/XPDA.git

[23] Sample online questionnaire can be found in: https://scsstcd.qualtrics.com/jfe/form/SV_2689YhiMINJGrpb